

Wills.
Ann Arbor
1925.



Digitized by the Internet Archive
in 2026

Why Children Succeed

A Study of the Factors Conditioning the Progress of
Children in School, and of the Problems That Must
Be Solved before the Relationships between the Fac-
tors and School Success Can Be Reliably Determined

by

STUART APPLETON COURTIS

*Director of Instruction and Dean of Teachers College
Detroit Public Schools*

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the
University of Michigan
May 1925

Published by
COURTIS STANDARD TESTS
1807 EAST GRAND BOULEVARD, DETROIT, MICH.
Friesema Bros. Printing Co. Press

173
Copyright, 1925, by the Curtis Standard Tests.
1807 East Grand Boulevard
Detroit, Michigan



Printed in U. S. A.

Additional copies of this book may be secured by writing
to the address above. The price is \$1.50.

ACKNOWLEDGMENTS

The measurements and tabulations discussed in this dissertation were made possible by a grant from the Commonwealth Fund through the University of Michigan. They constitute a unitary portion of more extended investigations and experiments which are still being carried forward. However, it is a pleasure to have, at this time, the opportunity to express appreciation for the assistance of Dean A. S. Whitney of the School of Education, through whose efforts the acceptance of the grant was brought about, and of Frank Cody, Superintendent of Schools, Detroit, Michigan, whose permission to use the experimental schools connected with Detroit Teachers College provided a suitable field for active operation.

The study herein reported is itself an outgrowth of a similar, but less complete and extended analysis, of the measurements of classroom products made in the Gary Survey. That investigation was carried on under the direction of Abraham Flexner of the General Education Board, to whom, perhaps, more than to any other one person, the author is indebted for assistance in securing the grant for the present study. Grateful acknowledgment is also made for similar advice and assistance given by Professors Farrand, Thorndike, and Judd.

No study of this character can be completed except through the co-operation of many persons. Mention must be made especially of the fine spirit in which the principals, teachers, and children of the Marr, Wingert, and McGraw schools have responded to the many demands made upon them for tests, records, and special assistance. Also, acknowledgment is gratefully made for the assistance rendered by the following members of the Detroit Teachers College faculty: Miss Marie I. Rasey, head of the research department, director of the testing and tabulating program; Ernest W. McFarland, instructor in sociology, in charge of the sociological investigations; Miss

Theresa Maley, head of the health education department, supervisor of the physical measurements; E. G. Blackstone, instructor in administration; Walter G. Bergman, instructor in psychology, in charge of the statistical analysis and computations; to H. J. Kaufmann for assistance in making graphic records; to Paul T. Rankin, assistant director in charge of instructional research, and to Arthur B. Moehlman, director of statistics and reference, Detroit Public Schools, for assistance in planning the investigation and interpreting the results; to Professors Guy Montrose Whipple and Clifford Woody under whose direction and criticism the dissertation has been prepared; to Professor H. C. Carver for invaluable direction and assistance with the many statistical problems encountered; and to many others too numerous to mention whose assistance through tedious labor or by criticism and advice, has been no less important than it has been effective. The credit, if the contributions of this study should be found worthy of credit, must be shared with many, but for its errors and imperfections, the author assumes full responsibility.

S. A. COURTIS.

TABLE OF CONTENTS

Chapter	Page
ACKNOWLEDGMENTS.....	3
I THE PROBLEM.....	11
II HISTORICAL BACKGROUND OF THE PRESENT INVESTIGATION.....	21
III THE FACTOR CONCEPT: ITS NATURE AND IMPORTANCE.....	29
IV DEFINITIONS OF THE TERMS AND CONCEPTS USED IN THIS STUDY	37
V THE MEASUREMENT ACTIVITIES AND THE DETAILS OF PROCEDURE	53
VI JUSTIFICATION OF THE ADOPTION OF A NARROW DEFINITION OF SUCCESS.....	59
VII THE MEASUREMENT AND ANALYSIS OF PERFORMANCE.....	79
VIII THE MEASUREMENT AND ANALYSIS OF HEREDITY, MATURITY, AND TRAINING.....	161
IX SUMMARY AND GENERALIZATIONS.....	201

APPENDIX

I REASONS FOR THE CHOICE OF TESTS SELECTED	207
II SAMPLE TESTS AND RECORDS.....	211
III DETAILED INSTRUCTIONS FOR GIVING AND SCORING CERTAIN TESTS	231
IV DETERMINATION OF FORMULAE FOR EFFORT AND ADJUSTMENT FACTORS	243
V EQUIVALENCE OF PREDICTION FORMULAE	247
VI ILLUSTRATIVE COMPARISON OF THE CONVENTIONAL AND PIECE- MEAL METHOD OF DERIVATION OF PREDICTION FORMULAE.....	251
INDEX	261
REFERENCES: AUTHORS AND TITLES.....	268

TABLES OF RESULTS

Number	Title	Page
I	EVIDENCE OF EDUCATIONAL INEFFICIENCY.....	13
II	RELATION BETWEEN INTELLIGENCE AND SUCCESS.....	14
III	PROGRAM OF WORK.....	54
IV	AGE GRADE RELATIONSHIPS AMONG CHILDREN TESTED.....	56
V	DETERMINATION OF SUCCESS IN TERMS OF TEACHER JUDGMENT.....	62
VI	COMPARISON OF TEACHER'S ESTIMATE OF SUCCESS WITH AVERAGE MARK.....	63
VII	COMPARISON OF VARIOUS FACTORS WITH AVERAGE MARKS..	64
VIII	RELATION BETWEEN GRADE STATUS, AGE AT ENTRANCE, AND SCHOOL PROGRESS.....	66
IX	FURTHER COMPARISONS OF FACTORS WITH SUCCESS AS DETERMINED BY AVERAGE MARK.....	68
X	DISTRIBUTION OF AVERAGE MARKS OF 819 BOYS AND 790 GIRLS, GRADES 2-8, ON SCHOOL REPORT CARD AT THE END OF THE SEMESTER.....	69
XI	COMPARISON OF ABSOLUTE ABILITIES, AS MEASURED BY SCORES IN TESTS, OF BOYS AND GIRLS WHO RECEIVE THE SAME AVERAGE MARK.....	70
XII	RELATION OF MATURITY, SCHOOL MARK, AND ABSOLUTE ACHIEVEMENT IN STANFORD TEST FOR 178 BOYS RECEIVING AN AVERAGE MARK OF 2.5.....	72
XIII	RELATION OF ACCELERATION AND RETARDATION TO SCHOOL ACHIEVEMENT AS MEASURED BY THE STANFORD ACHIEVEMENT TEST.....	74
XIV	RELATION BETWEEN OBJECTIVE MEASURES OF SUCCESS AND VARIOUS FACTORS.....	76
XV	RELATION OF HEIGHT TO AGE.....	92

XVI	RELATION BETWEEN AGE AND SCORES IN VARIOUS PHYSICAL TESTS.....	93
XVII	COMPARISON OF VARIATIONS IN IMMEDIATE AND DELAYED RETESTING.....	100
XVIII	INDIVIDUAL VARIATION IN PERFORMANCE.....	111
XIX	RELIABILITY OF SCORING RATE AND QUALITY OF HANDWRITING.....	114
XX	VARIATION IN COMPUTED SCORES.....	117
XXI	STRAIGHT LINE ADJUSTMENT CURVES.....	119
XXII	RELATION BETWEEN QUALITY OF HANDWRITING, ANGLE OF SLANT, AND AMOUNT OF ADJUSTMENT.....	121
XXIII	SCORES OF INDIVIDUAL IN ELEVEN TESTS OF HANDWRITING WITHIN TEN DAYS.....	123
XXIV	SCORES OF FOUR INDIVIDUALS IN SAME TESTS.....	124
XXV	COMPARISON OF THE EFFECTS OF TWO STIMULI ON EFFORT..	127
XXVI	COMPARISON OF MAGNITUDES OF EFFECTS OF TEACHER'S APPEAL AND EXAMINER'S MONEY PAYMENT.....	128
XXVII	EFFECT OF EFFORT UPON CLASS ADJUSTMENT LINES.....	130
XXVIII	VARIATIONS IN PERFORMANCE IN PERFORMANCE SCALES....	145
XXIX	COMPARISON OF VARIATION IN SCORES IN RATE TESTS AND PERFORMANCE SCALES IN READING; GROUPS MATCHED IN STANFORD TESTS.....	151
XXX	COMPARISON OF VARIATION IN SCORES IN RATE TESTS AND PERFORMANCE SCALES; GROUPS MATCHED IN RATE TESTS	154
XXXI	CONSTANCY OF SCORES IN PERFORMANCE SCALES IN RATE TEST.....	157
XXXII	COMPARISON OF VARIABILITIES OF RATE AND PERFORMANCE TESTS.....	158
XXXIII	VARIATION IN SCORE REVEALED BY CORRELATIONS.....	159
XXXIV	VARIATION IN RELATIVE MATURITY IN MOTOR AND THOUGHT PROCESSES.....	163
XXXV	MATURATION OF THE PECKING INSTINCT IN CHICKS.....	165
XXXVI	VARIATION OF HEIGHT AND WEIGHT WITH CHANGE IN CONDITIONS.....	169
XXXVII	COMPARISON OF THE EFFECTS OF SCARCITY OF FOOD UPON CHILDREN OF DIFFERENT AGES.....	170

XXXVIII	DETERMINATION OF EFFECTS OF TRAINING UPON INTELLIGENCE SCORES.....	172
XXXIX	RELATION BETWEEN TRUE VOLUME, OBSERVED VOLUME, TEMPERATURE AND PRESSURE.....	174
XL	ERRORS IN WEIGHT PREDICTED FROM HEIGHT AND WEIGHT.	180
XLI	VARIATION IN COMPUTED BUILD FACTOR.....	183
XLII	VARIATION IN AGE-HEIGHT RATIOS.....	185
XLIII	CORRELATION OF PHYSICAL TRAITS WITH EDUCATIONAL ACHIEVEMENT.....	186
XLIV	ACCURACY OF PREDICTION OF STANFORD SCORES.....	193
XLV	ACCURACY OF PREDICITON OF HANDWRITING SCORES.....	198
XLVI	ASSUMED DATA FOR AGE AND GRADE	252
XLVII	VALIDATION OF FORMULA BY COMPARISON OF PREDICTED AND ACTUAL GRADE	252
XLVIII	SCATTER DIAGRAM OF AGE AND GRADE DEVIATIONS.....	254
XLIX	MAGNITUDE OF VARIATIONS CAUSED BY VARIOUS ATTENUATING FACTORS	255
L	CORRELATION BETWEEN GRADE, AGE, AND AGE AT ENTRANCE	259

ILLUSTRATIONS AND GRAPHS

Figure	Title	Page
1	RELATION BETWEEN INTELLIGENCE AND SUCCESS.....	5
2	RELATION BETWEEN SEVEN MAJOR FACTORS.....	50
3	COMPARISON OF RELATIVE COMPLEXITY BY MEANS OF CORRELATION SURFACES.....	93
4	INDIVIDUAL VARIATION IN PERFORMANCE.....	112
5	STRAIGHT LINE ADJUSTMENT CURVES.....	120
6	RELATION OF ANGLE OF SLANT AND AMOUNT OF ADJUSTMENT TO QUALITY OF WRITING.....	122
7	SCORES OF ONE INDIVIDUAL IN ELEVEN TESTS IN HANDWRITING....	123
8	SCORES OF FOUR INDIVIDUALS IN THE SAME TESTS.....	125
9	INDIVIDUAL VARIATION IN RESPONSE TO A MONEY STIMULUS.....	129
10	DISPLACEMENT OF GROUP ADJUSTMENT LINES UNDER EFFORT.....	131
11	FIELD OF VARIATION FOR ONE INDIVIDUAL.....	132
12	INDIVIDUAL VARIATION IN HANDWRITING TESTS ON DIFFERENT DAYS	134
13	VARIATION IN PERFORMANCE IN ADDING.....	135
14	VARIATION IN SCORE IN COMPOSITION TESTS.....	136
15	VARIATION IN RATE AND QUALITY IN TYPEWRITING.....	137
16	ILLUSTRATIONS OF LACK OF COMPARABILITY IN HANDWRITING SCORES	138
17	ILLUSTRATIVE SELECTIONS FROM A PERFORMANCE SCALE IN SPELLING	141
18	VARIATION IN PERFORMANCE IN A PERFORMANCE SCALE IN SPELLING	141
19	VARIATIONS IN SPELLING CONTROL.....	142
20	PORTION OF TEST ILLUSTRATING DIFFERENCE IN COMPLEXITY IN TEST ELEMENTS.....	149
21	COMPARISON OF PERFORMANCE IN SITUATIONS OF DIFFERENT DEGREES OF COMPLEXITY.....	150
22	MATURATION OF PECKING INSTINCT.....	166
23	EFFECT OF SCARCITY OF FOOD UPON HEIGHT AND WEIGHT OF BOYS.	169

24	EFFECT OF SCARCITY OF FOOD UPON HEIGHTS AND WEIGHTS OF CHILDREN OF DIFFERENT AGES.....	171
25	EFFECT OF HEREDITY UPON HEIGHT.....	180
26	EXTREMES OF DAVENPORT SCALE FOR BUILD.....	181
27	ILLUSTRATION OF USE OF PREDICTIVE FORMULAE IN STUDY OF INDIVIDUALS.....	188
28	HEREDITARY DIFFERENCES IN MATURATION OF INTELLIGENCE.....	189
29	VARIATION IN SUCCESSIVE DETERMINATION OF I. Q's.....	195

WHY CHILDREN SUCCEED

CHAPTER I

The Problem

Education an American Ideal. Universal free public education is a distinctive American ideal. While some provisions for the education of the young are to be found among all races and peoples, yet the clear recognition of education as essential to the growth of democracy has naturally developed most rapidly in a country in which government from the beginning has been built upon democratic ideals.

The careless reader of history might be tempted to deny education a place among the early American ideals. Statements like the following are to be found.

"The Federal Constitution made no mention of any form of education for the people, nor does the subject occur in the debates of the Federal Constitution Convention."¹

Yet in 1776 nine colleges and a number of private and church elementary and secondary schools are proof that although the first educational efforts were distinctly voluntary on the part of the community, recognition of the importance of education and belief in the ideal of making adequate provision for the education of the young, were operative during the colonial period.

Present and Future Significance. The history of public education in the United States has been a story of rapid development and integration. Today, (1923) in the 271,319 schools in the country, more than 21,578,000 children are taught by some 679,533 teachers, and the cost of maintaining this great system

¹Cubberley, E. P. *Public School Administration* (1916) Chapter One, page 3.

is \$1,036,151,209.² Today the battle is being waged to establish a national bureau of education and future years seem destined to witness the further integration of our schools, this time on a national basis. Moreover, philosophers, sociologists, political parties, and many other agencies which influence public opinion have come to see and to preach that education in its broadest sense is the means by which the continuity of our civilization is maintained;³ and that because of this fact, education is the basic means of social control of human evolution and national destiny.⁴ The recent rapid multiplication of private agencies for the continued training of adults seems to imply that the development of education will be greatly extended in the future.⁵

Present Inefficiency. Because of the high value placed on education by the American people, whenever they have turned an appraising eye upon their schools, their criticisms have been many and severe. Whether we consider the comments made by the first American Surveyors,⁶ or current attacks on public education⁷ the essential content of criticism is always the same; the very great failure of the schools to perform their functions efficiently. It matters not that the improvements in education have been many and far reaching,⁸ or that every attempt to compare the past with the present on the basis of direct measurement has proved that present-day schools teach more efficiently than those of former days.⁹ The crux of the situation in education has been, and is, that the results of school training are less in amount and more uncertain in quality than the public of any age believes is justified.

How great the inefficiency of public education really is few realize in spite of the repeated revelations of survey data. The results of Rice,⁹ the pioneer in the modern movement for exact comparative measurement, were received with open disbelief

²Bulletin, Bureau of Education (1923).

³Dewey, John. *Democracy and Education*, Chapter I (1920); Cooley, C. H. *Social Process* (1922), page 76.

⁴Ward, H. T. *The New Social Order* (1919), pages 304, 428.

⁵Caldwell, O. W. and Curtis, S. A. *Then and Now in Education* (1923), page 59, Chapter V.

⁶Pritchett, H. S. *Seventeenth Annual Report of Carnegie Foundation for the Advancement of Teaching* (1922), pages 93-117.

⁷Smith, E. R. *Education Moves Ahead* (1924), Chapters II, III.

⁸Riley, J. L. *The Springfield Tests* (1908); Tirrell, H. A., "The Norwich Tests," *School Review* (1910), pages 326-332.

⁹Rice, J. M. *Scientific Management in Education* (1913).

and ridicule; but they have been substantiated by survey after survey the country over. Today there can be no excuse for ignoring the fact that very few children profit as they should from training experience in any single function, and that the range of ability in any one grade is so great that each grade is practically an ungraded group with respect to a single specific ability.

Evidence. In Richmond, Virginia,¹⁰ the result of training in reading upon the children in the seventh (and last) grade of the elementary schools as measured by the Thorndike Silent Reading Test Scale Alpha 2 is given as:

TABLE I
Evidence of Educational Inefficiency
Distribution of Individual Scores in Silent Reading

Thorndike Scale 2—Grade 7, Four Schools
127 White Pupils

Scores					Total	Median	Grade Interval	Grade Range
5	6	7	8	9				
1	49	59	17	1	127	7.2	.6	3.7

To interpret this range of ability one needs to know that the standard for the fourth grade is 5.2 and for the seventh grade 7.0, so that the average progress from grade to grade is a change in score of .6. As the scores in the seventh grade range from 5.9 to 9.6, or 3.7 points, the variability shown in Table I represents a range more than six times the average difference from grade to grade. In Richmond, elimination is given in the survey report (p. 40) as the reason for the smaller range in the seventh grade; for in the fifth grade the range is from a score of 2.2 to 8.1 or nearly ten times the average grade interval. Similar individual variation has been found whenever surveys have reported the distribution of the individual scores.¹¹

¹⁰Virginia Public Schools, Part II, *Educational Tests* (1921) page 37, Table 18.

¹¹Stone, C. W. *Arithmetical Abilities* (1908) page 74. Courtis, S. A. *Final Report, Committee of School Inquiry* (1913) Part II, pages 389-546. *Survey Report Board of Education*, Salt Lake City (1915). Courtis, S. A. *Measurement of Classroom Products*. General Education Board (1916) and many other surveys.

Data from the Virginia survey have been presented because it is a recent survey and gives also the scores of seventh grade city children in Virginia in an intelligence test. In former days schools were expected to turn out a uniform product; that is, a graduate of an elementary school was expected to have attained a certain level of ability. Consequently the revelation by measurement that such graduates vary widely in ability was a great shock. The development of tests of general intelligence, however, has led to further light on the situation. Intelligence tests are supposed to measure primarily, not the results of specific training, but rather native capacity.

For instance, the results of intelligence tests in the Virginia schools adjusted in frequency to a total of 127 (Table II) show a range of score from 35 to 145, or 110 points. This is also six times the average grade interval. ($110 \div 16$).

TABLE II

Relation Between Intelligence and Success

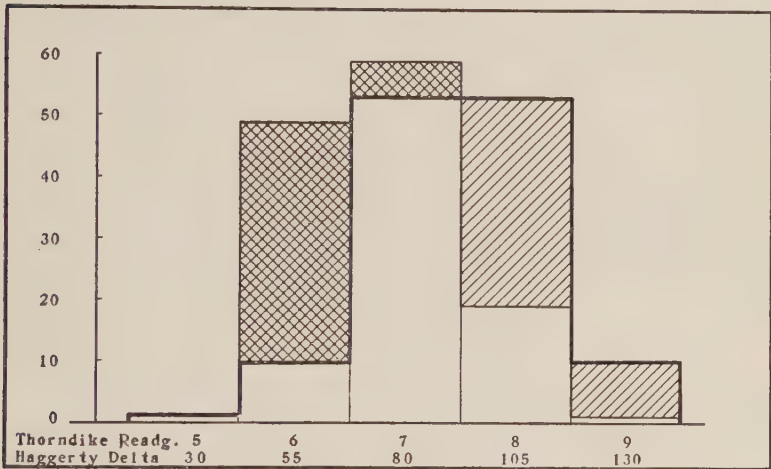
General Intelligence, Examination Delta, 127 City White Pupils from Nine Virginia Cities (Table 74 adjusted on a percentage basis from 728 to 127, also Table I repeated for comparison.)

Intelligence Score	30	55	80	105	130	Total	Median	Average Grade Interval	Grade Range
Frequency	1	10	53	53	10	127	105	16	110
Table I Frequency	1	49	59	17	1	127	7.2	.6	3.7
Reading Score	5	6	7	8	9	—	—	—	—

If these data are compared with those in Table I (as in the lower half of Table II) the comparison suggests that success in school work is conditioned by intelligence since the two distributions are so much alike. They suggest, also, that the schools are not training the pupils to the limit of their capacities. If they were, the two sets of frequencies would correspond; whereas the reading is high in the lower values and low in the higher values (Fig. 1).

FIGURE 1

Educational Inefficiency as Revealed by a Comparison of
Relative Development of Intelligence and of Ability in Reading



The surface of frequency enclosed by the heavy line represents a distribution of pupils on the basis of intelligence. The second surface of frequency represents the distribution of the scores in reading of approximately the same pupils. The shaded areas make clear the extent to which the two distributions do not coincide. If every child were trained to the limit of his capacity, the shaded areas would disappear. The shaded area, therefore, represents inefficiency.

Interpretation. These data, while valueless as scientific evidence, portray strikingly, but simply, the situation in education. Through the accumulation of scientific knowledge about the nature and results of the educational process and about the children who are affected by it, it has gradually come to be recognized by students of education that teaching and learning are both much more complex processes than had been supposed; that the final outcome is affected by many factors of which the efforts made by pupil and teacher are but two; that before there can be any true evaluation of the efficiency of teaching effort, any scientific control of the educational process, there must be a determination of the factors conditioning the progress of children in school and of the laws governing their effects.

In other words, it seems as if education has reached a stage comparable to the level of development attained in the physical

sciences toward the end of the eighteenth century. At that time, as a result of the application of scientific methods to a new field, the concepts of the alchemists and mystics of the middle ages were breaking down and new techniques were developing. To the fundamental elements of the ancients—fire, water, earth, and air—and the three principles—salt, sulphur, and mercury—phlogiston had been added in an attempt to adjust the old scheme of things to the newly acquired information. Nevertheless, conflicting results, unrest and endless discussions continued and were generally prevalent. Then came the precise studies of Priestley, Scheele, Lavoisier, Cavendish, Boyle, John Dalton, and others, yielding new modes of analysis and relatively precise statements of underlying natural laws.¹²

The industrial revolution which followed a century later is an expression of the increase in human power which flows from discovery and mastery of nature's laws. In other words, progress takes place whenever it is possible to analyze a complex situation into factors, to define each one explicitly, and to determine the laws of their influence.

The Present Need in Education. The time seems ripe for an attempt to prepare the way for a similar development in education. There is need to array the evidence that proves the existence of distinctive factors, to formulate definitions of essential concepts, and to attempt the experimental evaluation of the relations between such factors.

Up to this time education has been concerned chiefly with pooled results and massed data. These are, of course, of service in determining general tendencies. General tendencies are, however, but a part of the story. The goal is prediction and control of the behavior of individual children. This is likely to be secured in education, as elsewhere, only through the use of particularized data made comparable through correction for the effects of divergent factors.

Range of Problems Studied. This dissertation rests on the belief that ultimate success in educational science will be

¹²Williams, H. S. *History of Science*, (1904) Vol. IV, Chapters I and II.

achieved only when more attention than at present is paid to objective proof of the validity of the concepts, to determination of the precision of the instruments and processes of measurements, and to a study of the adequacy of the statistical methods employed.

The task, however, is both large and difficult. In the nature of things, little more can be done at present than (1) formulate an organized statement of the problem; (2) indicate the general methods of solution; (3) present supporting data from existing sources as well as from a new series of experimental studies, and (4) map out a program of problems or experimental investigations which must be completed successfully before an ultimate solution of the problem can be attained.

However, there is a real need for a systematic collection of basic facts and for an orderly presentation of these facts arrayed in proper perspective to form a picture of the problem as a whole. The distinctive contribution of this dissertation is such a picture. In addition there will be given some suggestion of developing techniques adapted to the task that lies ahead, and certain generalizations or concepts which may prove generally serviceable. In the main, however, contributions of this dissertation are of the problem-setting and not of the problem-solving type.

General Statement of Aim. More specifically, the aim of this dissertation is two-fold in its nature. One purpose is to isolate and define more completely than at present the factors that determine a child's success in school, or, where isolation of factors is not completely possible at present, to formulate the experimental or statistical investigations that must be undertaken to bring final success within reach. Here the goal is the determination of how reliably the success of a child can be predicted from the measurement of a few basic factors, with existing tests and scales.

The other phase is the philosophic purpose of utilizing the resulting knowledge to interpret the scientific study of educational problems, particularly the use of measurement in educational investigations, in such a way that both its value and its

dangers may be clearly seen. The research student in any field has frequent need to maintain his perspective, to weigh values and appraise proportions. A little knowledge is always a dangerous thing and especially during periods of rapid development is there need for frequent periods of self-orientation. Here the goal is the formation of a program of future work, a map of the road to be followed: its twistings and turnings, its difficulties, its ultimate destination.

The two aspects of the fundamental aim are inseparably interwoven throughout the thesis. Certain chapters may for a time seem to emphasize one aspect more than the other, but when theory is discussed, it is for the purpose of interpreting practice, and when experimental results are presented it is for the purpose of developing theory. From the point of view of purpose and need, therefore, the real problem of the dissertation is a single problem in spite of its dual nature.

The most serious difficulty in presentation is the danger of obscuring the real problem in the presentation of details. As has just been indicated, the basic purposes are orientation and generalization. But generalization is possible only from particulars, and particulars never represent more than a few of the characteristics of the general.

Program of Work. Thus, the first step in the solution of the problem was to define "success" and set up objective criteria for its evaluation.

The second step was to measure the success of children in terms of these criteria, and to measure also such other characteristics of the children as were thought to be related to success.

The third step was to canvass the data for the purpose of discovering the gross factors operating to produce success, to define these factors objectively, to organize the data which prove the existence of such factors, to determine the relative contribution of these factors to success and to attempt to express fundamental relationships, where such were discovered, in formulae which would aid a person in predicting a child's success from appropriate information about essential factors.

The fourth step was to predict, by means of these formulae, the scores made by children in the same tests given a year later, and to compare predictions with actual scores in order to measure the reliability of the formulae.

The fifth step was to appraise critically the conclusions reached and to formulate the specific studies and investigations that must be completed before the conclusions may be established on a more reliable foundation.

Limitations. Stated in these terms, the program is seen to conform to the problem. But add that success was defined in terms of score in the Stanford Achievement Tests and in handwriting tests, and the narrowness of the field obscures the purposes for which the investigation was undertaken.

Therefore, the adoption for this study of objective measures of success in attaining certain of the simpler and probably less important products of educational effort should be looked upon merely as a first, but essential, step towards the solution of the larger problem. There is probably no more certain road to eventual control of the more complex products of the educational process—reasoning power, ideals, character—than through the discovery of the laws which make possible control of the lesser products of school work. While the development of honesty may be a process very different from the development of ability to write, yet if both take place under law, the discovery of the factors and laws that condition the latter will inevitably stimulate similar investigations of the former, and methods of analyses developed for the simpler will apply, in part at least, to the more complex field.

Specific Statement of the Aim. The specific aim of this thesis is thus a study of the measurement of ability in the Stanford Achievement Tests and in handwriting with a view to determining both the factors and laws involved and the methods by which such measurements and analyses may be improved.

Relation of Specific and General Aims. The relation of the specific and general aims may be seen by a review of the form of presentation which the thesis follows. It begins with an

historical summary of the movements which have contributed to the present point of view, a statement of tendencies in the development of theory. Next, an explanation of the factor concept is given, and a summary of definitions of essential terms. Both of these discussions are present day products of past development of theory.

The discussion then turns from the theoretical aspects to the details of the actual measurements made, and the conditions under which they were made. The first fruit of these measurements, however, was evidence which justified the pursuit of the general aim in a narrow field. That justification is given in some detail.

So far all has been introduction and preparation. With the ground thus cleared and the problem rendered specific and definite, two chapters follow dealing in narrow and intensive terms with the measurement of the four major factors, performance, heredity, maturity and training in the specific fields selected for study.

Finally, there is a return to the general problem and the conclusions from the specific studies are utilized in making the larger generalizations which form the real goal of the study as a whole.

If this general plan is kept in mind, no difficulty will be experienced in passing and repassing from theory to practice and from specific to general as the needs of the fundamental purpose require. Also, the essential unity of the problem which the thesis as a whole discusses, will be apparent.

CHAPTER II

Historical Background of the Present Investigation

Aim. The purpose of this thesis is synthesis, its method is analysis. Such an investigation is never undertaken except where it has been preceded by many investigations differing widely in methods, character, and products, each contributing to a growing store of knowledge, but contributing independently. Yet whenever one pauses in the midst of any life activity and attempts to formulate the history of the influences which have led up to that activity, one at once becomes aware of the complexity of the web of life. It is difficult to trace in isolation even a single strand. Any historical statement must, perforce, be partial, a generalization made possible only through ignoring inter-relations and connections with other influences. Nevertheless, such imperfect generalizations serve a purpose in that they make clear the evolutionary tendencies at work and give the viewpoint from which the next forward step is to be interpreted. This chapter undertakes to present in orderly sequence the movements and developments which have made this study possible.

Periods. Man has always been interested in the workings of his own mind, but from the speculative, fanciful explanations of the ancients to the psychological science of the present day is, indeed, a far cry. For the purpose of this study, however, the evolutionary process in the field of inquiry with which this study is concerned may be divided roughly into three periods: Descriptive psychology (from the Greeks to approximately 1875), experimental psychology (from 1875 to 1900), and educational psychology (from 1900 to the present).

Descriptive Psychology. The psychology of the ancients is called descriptive because it was speculative and metaphysical in nature concerned chiefly with describing, classifying and naming mental phenomena. It was almost wholly deductive and some of the explanations which it evolved were fanciful in the extreme. In general the soul of man was conceived as being fitted out with a number of ready-made faculties, memory, will, attention, etc., which were thought of as separate limbs or members of the mind. The faculties, originally, were distinguished not only from each other, but from the essence of mind.¹ Faculty psychology dominated educational thinking during the middle ages and many traces of its influence can be found in the current educational thought and practice.

Experimental Psychology. The period from 1750 to 1850 was one of remarkable scientific development in many fields. In chemistry, physics, physiology, and kindred sciences, the application of inductive methods was yielding epoch-making discoveries. The nerve physiologists (Bell, Desmoulins, Hall, Bernard) by application of the same methods, soon after gave the world an accurate, if meager, knowledge of the anatomy and functions of man's nervous system. In similar fashion the psycho-physicists (Helmholtz, Weber, Fechner), applying the methods of exact measurement to psychological activities, paved the way for the union of all the developing techniques and discoveries into that new phase in the history of psychology designated above as experimental.²

Sir Francis Galton in England,³ and Wundt in Germany,⁴ most often come to mind as representing the beginning of modern experimental psychology. Each at about the period represented by the date 1875, began to apply scientific methods to the study of the nature of man and of the effect of environment upon that nature. The outstanding contributions of these new methods and fields of investigation were (1) measuring instruments, or psychological tests, (2) testing and statistical procedures, and (3) a very

¹Hamilton, Sir W. Discussion on Philosophy, Education and University Reform, p. 272. Quoted in Monroe's *Encyclopedia of Education*.

²Williams, H. S. *History of Science*, Vol. IV. Chapter IX (1904).

³Galton, Sir Francis. *Hereditary Genius* (1869).

⁴Wundt, W. *Outlines of Psychology* (1897).

great increase in our knowledge of the nature and development of man, particularly of individual differences. In this country Cattell and his students⁵ were the pioneers in a field which has expanded very rapidly. Whipple's *Manual of Mental and Physical Tests*, published in 1910, a book of directions compiled with special reference to the experimental study of children in the laboratory or classroom, furnishes an admirable summary of the contributions of this phase of psychology to education in tests, results, and conclusions.

Educational Psychology. Psychological measurements furnished data in conflict with the explanations of behavior given by the faculty theory. The attention of the experimental psychologists was soon being directed to a study of the learning processes in animals as well as man. Toward the close of the nineteenth century, Rice, inspired by the work of the German experimentalists, hit upon the comparison of school systems by giving them examinations under identical conditions. In the person of Thorndike the psychological and educational aspects fused. The educational test and scientific methods of test and scale construction are significant products of that union.

In the meantime developments in industry emphasizing efficiency and in sociology emphasizing the social traits and relations of man had also been taking place. In education all these tendencies found expression in the school survey movement (1910 to 1920) which popularized the use of educational measurements and added greatly to our knowledge of the effects of the educational process. The third phase of psychological development, and the one upon which this study is directly based, educational psychology, may thus be said to have begun about 1900.

Measurement of Physical Characteristics. During the period from 1875 to the present another, and in a measure quite independent though related, development has taken place, the measurement of human differences in growth and capacity, both physical and mental. In *Hereditary Genius* Galton discussed an imaginary scale for the measurement of general ability, basing the scale upon the theory of a normal distribution and giving it

⁵Cattell, J. McKeen. "Mental Tests and Measurements," *Mind*. Vol. XV, (1896) pages 373-380.

fourteen steps. Galton raised to prominence the question of the relative effect of nature and nurture, and for himself, chose nature as the basic factor to be first controlled. The founding of his Anthropometric Laboratory in 1884, and of the Eugenics Laboratory in 1905, are expressions of his belief that the fundamental step in the solution of many of our social problems is to make improvements in our racial stock through care in selection and breeding of natures of great innate capacity. As a result the English scientists since that time have paid great attention to measurement of physical and mental development. The outstanding men from whom, directly or through their pupils, many contributions have come are Pearson and Spearman.

In America, we find Bowditch⁶ measuring the physical growth of school children and using Galton's newly devised statistical methods as an aid in interpreting his results. Boas, Porter, Hastings, Wood, and others have greatly extended the range and precision of these measurements. Today Baldwin⁷ and others are carrying on anthropometric measurement of children in a very thoroughgoing manner and are giving to investigations of this type a new turn by measuring the same children repeatedly during their growth periods and studying the relation between individual curves of physical and mental development.

Measurement of Intelligence. Another line of development which has made very important contributions to the present study had its beginning in the care of handicapped children. An institute for deaf mutes was established in Paris, in 1712, for the blind about 1798, for the feeble-minded in 1837. In America special classes for backward children were organized in public schools about 1893. Very soon after the beginning of the twentieth century, extreme defects in mental capacity had been recognized as a special educational problem and psychologists were already beginning to work on the problem scientifically.⁸

Measurement of intelligence, as we know it today, has its

⁶Bowditch, H. P. "The Growth of Children." *Tenth Annual Report*, (1879) State Board of Health, Mass.

⁷Baldwin, B. T. "Physical Growth of Children from Birth to Maturity." *University of Iowa Studies in Child Welfare*, No. 1 (1921).

⁸Norsworthy, N. "The Psychology of Mentally Deficient Children." *Archives of Psychology*, No. 1 (1906).

Witmer, L. "Clinical Psychology." *The Psychological Clinic*, Vol. 1, No. 1 (1907).

immediate source in the work of Binet, Director of the Laboratory of Physiological Psychology at the Sorbonne. His set of tests was first published in 1905, and revised in 1908 and 1911.⁹ One of Binet's contributions was the distinct formulation of the concept of mental age as different from chronological age.

The Binet scale was introduced to American psychologists and educationalists through the work of Goddard,¹⁰ Director of the Psychological Laboratory at the Training School for Feeble-minded Children at Vineland, N. J. While the original purpose of mental testing was the selection of feeble-minded children, out of the criticisms and discussions which arose, the idea of the general measurement of intelligence gradually emerged. In the meantime the tests and the testing procedure were undergoing revision and improvement. The publication by Terman¹¹ of the Stanford Revision of the Binet scale, making use of the Intelligence Quotient, previously suggested by Stern,¹² made available for general use both a practical instrument and a fundamental concept.

The time required for individual testing and the widespread use of group psychological and educational tests naturally led to the attempt to construct group tests of intelligence. This advance was under way at the entry of the United States into the World War, but received enormous impetus and publicity through the use of group intelligence tests by the army psychologists.¹³ Today the use of group tests of intelligence for classification of children is becoming a routine part of school procedure.

The recognition of the individual differences in the capacities of children led inevitably to evaluation of school achievement in terms of capacity. The concepts of Educational Age, Educational Quotient, and Accomplishment Quotient, are the latest contributions of the development which it is the function of this chapter to record.¹⁴

⁹Binet, A. (1895-1911) Many articles in *L'Année Psychologique*, particularly *Le développement de l'intelligence chez les enfants*, Vol. XIV, pages 1-94.

¹⁰Goddard, H. H. "The Binet-Simon Measuring Scale for Intelligence." *Training School Bulletin*, January (1910).

¹¹Terman, L. M. *The Measurement of Intelligence* (1916).

¹²Stern, W. *Die Intelligenz der Kinder und Jugendlichen*, page 143, (1920).

¹³Yoakum, C. S. and Yerkes, R. M. *Army Mental Tests* (1920).

¹⁴Franzen, R. "The Accomplishment Quotient," *T. C. Record*, Vol. XXI, (1920) pages 432-442.

Appraisal. If now we attempt to appraise the results of these various types of activities and developments in terms of enduring scientific laws from which alone control of the educational process can come, the net product appears very small indeed. The world has gained a vast fund of miscellaneous, unorganized items of knowledge about human nature and its development; it has formulated quite clearly a number of fundamental concepts, and a few general laws. But of a science of education comparable with the sciences of physics and chemistry, it must be admitted that only the materials have been gathered. The basic framework of its foundation has yet to be constructed. As Spearman says:¹⁵

"The experimental growth is still in its infancy; and it appears to suffer from some pernicious infantile malady. Its results, dazzlingly important though they may sometimes be, are nevertheless astonishingly exiguous in proportion to the enormous outlay of labor upon them. Even at their best, they always fall most disappointingly short of what had seemed to be their legitimate expectations. Their triumphs attain to the dimensions of **splendid raids**, rather than to those of **consolidated conquests**."

Spearman quotes approvingly the estimate of James¹⁶ to similar effect:

"A string of raw facts; a little gossip and wrangle about opinions; a little classification and generalization on the mere descriptive level; a strong prejudice that we have states of mind and that our brain conditions them; but not a single law in the sense in which physics shows us laws, not a single proposition from which any consequence can casually be deduced. We don't even know the terms between which the elementary laws would obtain if we had them. This is no science, it is only the hope of a science."

In view of this long history, and of the difficulties in the way, it should be evident that little ought to be expected from such investigations as those of this thesis except movement in the right direction. The specific contributions of this dissertation may be considered either small or large; small, if measured in terms of facts and principles which may be accepted as universally established; large, if, as the author hopes, it provides problems,

¹⁵Spearman, C. *The Nature of Intelligence*, (1923) page 28.

¹⁶James, W. *Psychology, Briefer Course*, (1892) page 468.

facts, and techniques which stand the test of use by others, and which aid in the universal struggle for mastery, through scientific knowledge, of that most important means of social control, the education of the oncoming generation.

CHAPTER III

The Factor Concept: Its Nature and Its Importance

Natural Law. This dissertation is built upon a belief in the potency of analytical processes to reveal simple unitary elements or forces in the complex entities which present themselves to the eye and mind of natural man. The factor concept is one of the cornerstones upon which the whole structure we call science rests. A natural law is a statement of the relation between factors and results considered as causes and effects. It expresses in mathematical terms the influence of each factor in producing some complex result. These relations are determined in every case by experimentation which conforms to the law of the single variable. That is, in a complex situation, all the factors but one are kept constant, and the changes produced by variations in the one factor are measured. The generalized statement of these results is a law.

Usually a law is true only within limits because either (1) the influence treated as a factor includes the combined effect of many cooperating factor elements so that the exact relation expressed by the law holds true only as long as the factor elements act together; or (2) outside the limits continued action brings into play new factors which also contribute to the result. In other words, all human formulations of natural laws are relative; they are not formulations of casual relations based on a knowledge of fundamental relationships, but empirical generalizations from experiences. Nevertheless, as long as similar conditions prevail, knowledge of law is easily transmuted into control, within the limits of the law. The method of the inductive sciences is, therefore, essentially analytical, and progress is made whenever in a

complex situation, an apparent factor is isolated and its functional relationship to the general result determined.

Genesis. The fundamental assumption underlying the factor concept is that "under identical conditions identical causes lead to identical results." The concept of cause and effect arises early in the development both of a race and an individual. The first stage of consciousness is, in terms of the famous description by James,¹ "one big blooming buzzing Confusion." The first time a child sees the light, he *is* it rather than sees it, but as experiences multiply, the general sensory confusion becomes resolved into parts and in time he comes to perceive objects as space-occupying, moving things, external to himself. For a long time movements of objects take place under conditions which are mysterious and inexplicable, but gradually the concept of cause and effect makes its appearance and order begins to emerge from chaos.

In considering this evolution it is important to note that the factor concept arises by a selective process. Man tends to remember the constant elements of experience and to strip off, and to throw away as unimportant, the inconstant features. A factor, for this reason, is always an abstraction. Thus whether we consider material objects,² forms of energy,³ modes of behavior,⁴ or emotional states,⁵ the thing considered is always an abstraction and any definition of it is always partial. The philosophers in their persistent search for reality have made it very clear that there is no proof of the existence of anything apart from the consciousness which knows it.

"Our objects of knowledge are fettered by the forms and the limitations of human consciousness: they exist, as Kant has shown, under the subjective forms of space and time, of causality and the other relations."⁶

In other words, the universe is a single whole and the factor concept, which enables man to resolve it into parts, is a mere convention.

¹James, W. *Psychology, Briefer Course*, (1892) page 16.

²As rocks, trees, animals, etc.

³As heat, electricity, gravitation, etc.

⁴As honesty, truthfulness, idealism, etc.

⁵As fear, joy, sympathy, etc.

⁶Calkins, M. W. *The Persistent Problems of Philosophy*, (1907) page 365.

Reality is always larger, more inclusive, than the concepts by which it is apprehended. James⁷ says:

"The universe of all of us is still to a great extent such a Confusion (as the baby's), potentially resolvable, and demanding to be resolved, but not yet actually resolved, into parts. So far as it is unanalyzed and unresolved, we may be said to know it sensationally; but as far as parts are distinguished in it and we become aware of their relations, our knowledge becomes perceptual or even conceptual."

For practical purposes, however, the factor concept has great value. Indeed, the factors thus abstracted and given independent existence by the mind, as it were, have **so great natural value** that it is only the poet, the philosopher or the spiritual leader who, unaided, intuitively senses their underlying unity.

Dangers. For this reason there is grave danger that those who attempt to analyze human behavior by attempting to trace the effect of various factors will forget that in the last analysis the existence of the factors as separate entities has no more reality than is given to it by the process of abstraction. Consequently, there is need to realize at all times that the "factor convention" is a means to an end and not an end in itself. Otherwise formalism results.

For instance, the I. Q. is a convenient device for summarizing a series of data in regard to the behavior of a child. In the hands of those who keep in mind the limitations of this abstraction, the symbol is a serviceable tool. For those who forget it is an imperfect generalization of a part of the facts in the situation, it is accorded a respect which leads to educational determinism.⁸

On this point Cooley says:⁹

"The weak side of the standardization idea, as applied to society is its trend toward the numerical and mechanical. An external, visible test, almost always superficial in this sphere, is easy to apply and for that reason recommends itself to all who seek precise results without an exercise of the higher faculties of the mind. . . . To the statistical type of mind, precision is apt to seem in itself a guarantee of truth, and it is common to see elaborate calculations based

⁷James, W. *Psychology, Briefer Course*, (1892) page 16.

⁸Bagley, W. C. "Educational Determinism," *School and Society*, (1922) 15:373, April 8.

⁹Cooley, C. H. *The Social Process*, (1922) pages 385-388.

on assumptions which will not bear scrutiny. . . . Many accept numerical systems and precision as 'science' without further inquiry. . . . The statistical method does not require that the process we are testing be understood, but only that it be uniform. In that case its future working may be predicted from its past. There is a large and legitimate field for ingenuity in thus standardizing human function. Formalism is apt to come in, however, by our taking a mechanical view of the function itself, of the end to be sought, in order to make it more easily measurable."

Perhaps the best statement of the spirit in which the measurements herein reported are made is that of Thorndike:¹⁰

"Ultimately, indeed, every fact in human life is a case of the co-action of all the universe except the man in question, and the condition of the man in question at that instant. In taking anything short of all the universe save him and calling it the situation, we are abstracting—are replacing the total effective situation by some element of it. Also in taking anything short of the rich entirety of the man at that instant as the organism, we are abstracting—are replacing the total effective conditions of responses by some of their main features. Such abstraction is, of course, the procedure of common sense and of science. Everywhere there is abundant justification for building up an abstract scheme of responses which situation *a*, *b*, *c*, single evoke, though in fact they never act singly; or of bonds between situation *d* and a total set of responses, though in fact the component elements of *d* are never present in just the same proportions so that the very existence of *d* as a thing by itself is a myth."

Effects Not Additive. The attempt to deal with human behavior on the factor basis involves analysis of the products of conscious volition. The peculiar characteristic of consciousness is its synthesizing power—consciousness constantly creates and maintains a unity out of a diverse multiplicity. Apparently, however, unity is achieved not by an additive process, but by an exceedingly complex creative process. Wundt says:¹¹

"Wherever we may look around in the regions of those processes which we in the widest sense call psychic connections, or . . . since all actual mental processes are composite, and therefore, connections . . . wherever we may look around in the wide realm of psychic phenomena at all; always and everywhere the conspicuous characteristic meets us, that the product derived from any number

¹⁰Thorndike, E. L. *Educational Psychology*, (1913) Vol. I, pages 10-11.

¹¹Wundt, W. *Physiologische Psychologie*, (1911) Vol. III, page 555. Quoted by Spearman, *Nature of Intelligence and the Principles of Cognition*.

of elements is more than the mere sum of the elements, that it is more than a structure of the same kind as the elements deviating from them only in respect of quality or quantity; but that instead it is a **new event** which in its essential character is absolutely incomparable with the factors cooperative in producing it. This fundamental property of psychic occurrence we will call the principle of creative resultants."

This, and similar statements, are sometimes taken to mean that analysis of psychic action is impossible. In the judgment of the writer, however, it means nothing of the kind. He believes that the synthetic action of consciousness does not transcend the natural order; that under identical conditions even the creative process always produces identical results. A statement very similar to that of Wundt might be made about chemical action. Surely the hydrogen set free when zinc is added to dilute sulphuric acid is a new event, more than the mere sum of the elements, more even than a structure of the same kind as the elements, deviating from them only in respect of quality or quantity. Hydrogen, too, is incomparable in its essential characteristics with either zinc or sulphuric acid.

The difference in chemical and psychic situations would appear to be in their relative complexities. Psychic situations involve so many and such varied factors, and creative combinations take place under conditions so far removed from direct observation, that careful measurement and analysis of the actual process is much more difficult than in the case of chemical action. Nevertheless, the psychic process involves distinctive energies and the operation of differentiated structures and materials just as the chemical process does. Methods of analysis which have proved effective in one field should prove effective in the other to the degree that ingenuity is exercised in adapting them to the new conditions. The fundamental fact to be kept in mind is that already referred to above, that all human behavior without exception is a creative resultant and too complex to be accurately described as the product of any one factor.

Difficulties. There is one respect in which the application of the factor concept to the psychological field differs from its application to the physical sciences: that is in the ease with which

basic concepts are formed. Matter, time, and space are natural factors in that their perception as factors occurs before processes of generalization have been brought under conscious control, while concepts of capacity, mental age, ability, etc., arises only as the products of systematic generalization from much abstract reasoning. In the long run, however, both sets of concepts pass through the same process of progressive modification based on definition, measurement, criticism, and reformulation.

"The historical development of these concepts (space, time)," says one recent writer,¹² "culminating in the theory of relativity of Einstein, throws much light upon their nature. Perhaps the names of Euclid, Newton, Faraday, and Einstein, more than any others, are representative of the various stages in the evolution which has taken place."

Measurement in education is but fifteen years old at the most. Perhaps in the course of the next two thousand years, educational science may be able to boast a galaxy of names as great as those given above. It is not surprising, however, that now, at the very beginning of analytical study, our difficulties resemble those of the Egyptian surveyors on the banks of the Nile two thousand years ago as they wrestled with the problem of mastering the measurement of spatial relations.

It was precisely at this point of definition that the old faculty psychology went astray. Concentration, attention, memorization, and similar activities, were conceived of as unitary processes due to the operation of specific organs of mind, whereas any single act of concentration, attention, or memorization is as much a resultant complex as any other behavior. Indeed, to such a degree is this true that even if the basic activities of remembering, attending, willing, etc., were each carried on by a specific organ of the mind whatever the content, it is doubtful that our present technique of analysis would enable us to break up specific instances of behavior into their elements with sufficient precision to establish the existence of those organs.

The reaction from the faculty psychology has sometimes driven scientists too far in the opposite direction. They are in

¹²Birkhoff, G. D. "The Origin, Nature and Influence of Relativity." *The Scientific Monthly*, April (1924) XVIII:225.

danger of making the old mistake in a new form. As one writer puts it,¹³

"Nevertheless there is in current discussions a certain tendency to commit some of the fundamental fallacies of the earlier faculty psychology. The opponents of the doctrine of formal discipline, in their efforts to show that there is no such general training as has frequently been assumed, are subdividing consciousness into separate and distinct modes of activity, hardly less rigid and atomic than were the faculties. . . . The attack upon the general faculty of perception has thus led to a belief in many faculties of perception. These many faculties which might now be designated faculties for perception of squares, circles, triangles, etc., are separate and distinct in their character and in their training. Education must cultivate each in its order, we are told. . . . This modern form of faculty psychology is quite as objectionable as the older form."

Thorndike, discussing experiments on the spread of memory training says:¹⁴

"These experiments should at least put forever out of court the **two absurdities**—that improvement is in an absolutely **general 'memory'**; and that improvement is absolutely restricted to **exact and entire duplicates** of the situation met in the special practice."

Thorndike's answer to the questions, to what extent does improvement of one function alter others—and how—is that

"a change in one function alters any other only in so far as the two functions have as factors identical elements. By identical elements are meant mental processes which have the same cell action in the brain as their physical correlate."

Thorndike is here defining factor in the basic sense in which it will be used throughout this dissertation.

Definition. For the purpose of this study, however, the term, factor, will be used with a wider meaning. It will be applied (1) to any mental process, or organization of processes, sufficiently unitary to permit its contributions to some resultant complex to be recognized and evaluated;¹⁵ (2) to specific types of cell action in the nervous system which are the physical correlates of mental processes;¹⁶ (3) to any unitary influences altering the

¹³*Cyclopedia of Education*, Article on Faculty. (1911) Vol. II, page 570.

¹⁴Thorndike, E. L. *Educational Psychology*, (1913) Vol. II, page 397, italics mine.

¹⁵As the mental activity which determines spelling, writing, attending, etc.

¹⁶As the nerve cells whose action conditions the specific characteristics of spelling, writing, etc.

structure or condition of the cells which so act; ¹⁷ (4) to any unitary influences affecting the functioning of the cells;¹⁸ (5) to any elements affecting the processes of recording, tabulating or interpreting behavior.¹⁹

Purposes. There are two highly significant aspects of the present situation in educational science to whose explanation analysis by means of the factor concept may make important contributions. The first is the fact that correlations between human traits are in general positive. The second is the fact of individual differences. Many agreements and divergences in educational results may be completely explained when our knowledge of factors, and of the laws of their operation, are known.

¹⁷As training, growth, etc.

¹⁸As heat, drugs, fatigue, etc.

¹⁹As means, graphs, correlations, etc.

CHAPTER IV

Definition of Terms and Concepts Derived from the Past and Used in this Dissertation with Specific Meanings

The experimental investigations in the fields of psychology and education have yielded an enormous mass of data. They have revealed a number of fundamental characteristics of children and their abilities which are today generally accepted and which constitute the starting point of the analyses of this study. The more important of these are:

First, children of any one age group and sex differ greatly from each other, both in ability to profit by experience and in the experiences from which they profit.

Second, children of one age group differ from those of a younger age group in what is usually termed degree of development.

Third, experience modifies the characteristics of behavior of an individual in subsequent experiences.

Fourth, the number of characteristics, or aspects of behavior, in which children may be compared is endless so that comparisons must be made in terms of specific characteristics of behavior except as two or more, seemingly different, may be proved to be identical wholly or in part.

Fifth, it is not possible to compare characteristics or behaviors directly; the method of comparison is to measure certain quantitative aspects of a given characteristic or of behavior under set conditions involving the trait in question, and to call this measure of behavior a measure of the characteristic or trait itself.

Sixth, analysis consists of grouping the records of measurement until relationships between records are revealed.

There are, therefore, many opportunities for applying the factor concept in educational problems.

Repeated experience in analysis gives rise to concepts which create a need for terms. Analysis of success into factors is not possible without words to describe the different factors isolated. For the purposes of this investigation the major concepts will be taken as seven: heredity, maturity, training, ability, conditions, performance, statistical, and analytical procedures. Each will be defined in turn.

Heredity. The fundamental law of life is that all life arises from pre-existing life. While our ignorance of the process by which new lives come into being is still very great, such evidence as we have tends to show that spontaneous generation of life does not take place. Accordingly in each individual we have to deal with new combinations and re-organizations of existing biological elements. The possible combinations are infinite, and "individual difference" is the common characteristic of all living things.

In human reproduction two specialized cells, one from each parent, fuse to form a single cell. The life mechanism of this new cell is made up, by a process of selective combination, of elements derived from the reproductive cells. So far, then, as the offspring which results from the development of the new cell exhibits distinctive characteristics or behaviors which closely resemble those of a parent, he is said to have inherited the characteristic or trait. That is, heredity may be defined as "The resemblances among individuals due to their common origin or germ plasm."¹

A natural extension of meaning gives to the term a broader connotation. Heredity is sometimes used to differentiate causes which have their origin in the mitotic processes of fusion, selection and organization of biological elements from those which have their origin in the subsequent processes of development and living. In this sense, heredity means "determined by the content and fusion of the germ cells" without any reference to the resemblance of offspring to ancestors.

¹Cattell, J. McKeen. "An Interpretation of Intelligence Tests," *The Scientific Monthly*, May (1924) XVIII:509.

Early studies of heredity dealt chiefly with outstanding resemblances and a mechanical, deterministic explanation was gradually evolved that made of heredity something in itself, a force or principle or law that does things of itself. Later studies have revealed a process much greater in complexity, and very different in character, from that ordinarily connoted by the term heredity. The modern view, and one more in accord with our general knowledge of the workings of the universe, is that represented in the following statement made up of a number of quotations from a single article,² quotations which have been pieced together by me to present the view of heredity that will be upheld throughout the thesis.

"Organisms are like other objects in this respect; what they do or become depends both on what they are made of, and on the conditions surrounding them. . . . research has shown that the substances passed from parent to offspring, giving rise to the phenomena of inheritance, are a great number of discrete packets of diverse chemicals (chromosomes) imbedded in a less diversified mass of material (cytoplasm). The number of different kinds of packets that go into the beginning of any individual is very great, running into the hundreds or thousands. They are not massed in a haphazard way, but are arranged in a definite manner. . . .

"Development we know consists in (the) orderly interaction of these substances—with each other, with the rest of the cell body and with the oxygen, food, and other chemicals brought into the cell from outside; all under the influence of the physical agents of the environment. The final result—what the individual becomes—is dependent upon all these things; a change in any of them may change the result.

"The disposition of the chemical packets, or genes, is known to be at the beginning that of a double serial arrangement, like a pair of strings of beads. . . . When the organism becomes a parent, these sets of packets are distributed to the offspring according to a simple plan. . . . One parent gives to any particular offspring one packet only of each of its pairs. The other parent supplies the corresponding second packet of the pair, so that the offspring

²Jennings, H. S. "Heredity and Environment," *The Scientific Monthly*, September (1924) XIX:225. I have taken liberties with Professor Jennings' article. I have put together sentences, and parts of sentences from widely different portions of the article. I have supplied the connecting words necessary to make of the whole a smooth-running account; I have made the account present the view of heredity in which I believe. On the other hand, I consider that at no place have I misrepresented original meanings, and the whole is so exactly in the words of the original article that it must be called a quotation. Originality of ideas and clarity of expression, therefore, should be credited to Professor Jennings, while such misunderstandings, distortions, and mistakes as may appear when my abstract is compared with the article as written should be charged to me.

has again the full complement of pairs. . . . The laws of inheritance are not immediate consequences of some fundamental physiological principle, but of the arrangement of the packets of chemicals and their method of distribution. . . . as the rules work out in most cases of bi-parental inheritance, every germ cell gets a different combination of these packets of chemicals from that obtained by any other, so that in consequence every individual starts out as a different combination of chemicals from every other. . . .

"Substituting one or more packets for others is found to change the characteristics of the organism produced; different sets give when they develop, even under similar environments, different physical, mental, and moral peculiarities. . . . (It is this fact that has led to the idea of unit characters. It was thought that) there was one gene for eye color, another for stature, another for feeble-mindedness, and so on. . . . Advance in knowledge of genetics has demonstrated (the falsity of this idea). . . . Recent investigations (show that) the genes interact in complex ways for long periods; and every later characteristic is a long-deferred and indirect product of this interaction. Into the production of any characteristic has gone the activity of hundreds of the genes, if not all of them. . . . In the fruit fly, at least fifty genes are known to work together to produce so simple a feature as the red color of the eye, . . . and each of the cooperating packets is necessary; if any one of the fifty is altered, the red color of the eye is not produced.

"The genes then are simply chemicals that enter into a great number of complex reactions, the final upshot of which is to produce the completed body. The characters of the adult are no more present in the germ cells than is an automobile in the metallic ores out of which it is ultimately manufactured. To get the complete, normally acting organism, the proper materials are essential; but equally essential is it that they should interact properly with each other and with other things. **And the way they interact and what they produce depends upon the conditions.**

"Not only what a cell within the body shall become, but what the organism as a whole shall become, is determined not alone by the hereditary materials it contains but also by the conditions under which those materials operate. Under diverse conditions, the same set of genes will produce very diverse results. . . . It is not true that because an individual inherits the basis for a set of characteristics, that he must have those characteristics. . . . The same set of genes may produce many different results, depending on the conditions under which it operates. True it is that there are limits to this; **that from one set of genes under a given environment may come a result that no environment can produce**

from another set. But this is a matter of limitation, not of fixed and final determination; it leaves open many alternative paths . . . it is not necessary to have a characteristic merely because one inherits it. Or more properly, characteristics are not inherited at all; what one inherits is certain material that under certain conditions will produce a particular characteristic; if those conditions are not supplied, some other characteristic is produced. . . .

"The educated man has characteristics very diverse from those he would possess if uneducated. We say, when we think of this fact, that these are acquired characteristics, environmental characteristics, due to education. This is correct; but there is a tendency to go farther and say that these are not inherited characteristics, which is a mistake. The characteristics of the educated man are his native, inherited characters, just as truly as are any that he has. For all his characteristics depend on the conditions under which he develops, and would be diverse under different conditions, just as is true of the characteristics that develop under education. And the characters developed under education depend upon the hereditary materials derived from his parents, changing as these materials are altered, just as do all others. . . .

"In the group of organisms to which man belongs there is an early period in which it is practically difficult to change effectively the conditions under which the organism develops, because it is enclosed within the mother's body or within a resistant egg shell. So we have gotten accustomed to calling inherited those characteristics which are determined before it leaves its mother's body or the egg, while those determined later are called acquired characters. But this is an artificial distinction based on practical considerations. . . . anything the **environment brings out is hereditary.** All characteristics are hereditary and all are environmental. . . . No single thing that the organism does depends alone on heredity or alone on environment; always both have to be taken into account.

"The concepts of the hereditary and the environmental cannot be employed in the absolute way now practiced; but they can be used with entire precision if they are applied, **not to characteristics-in-themselves, but to the diversities between different particular cases.** Though stature is always dependent on both heredity and environment, the difference in stature between Mr. Jones and Mr. Smith may be purely a matter of heredity; the difference between the same Mr. Jones and Mr. Brown may be purely a matter of environment. If there is clarity as to what comparison is made, there need be no ambiguity as to what is due to heredity, what to environment."

In other words there are two series of causes whose interaction determines each educational product; one of these, consisting of matter and organizations of matter, is derived from parents; the second, consisting of the play of matter and energy upon the first, comes from the environment. To name either apart from the other is to make an abstraction, is to set up a distinction which is partial and relative, not absolute. The use of the term, heredity, is therefore, a signal to the reader that for the moment attention is to be directed to one aspect of a situation to the neglect of the other, and vice versa, for environment. In general, during the period before birth in which environmental forces are not under control, and appear to be almost constant from generation to generation, the cause of differences will be designated as heredity. After birth, when environmental factors are deliberately manipulated to produce desired results the resulting differences will be considered environmental. Also, at any time, when environment forces are held constant, and differences in effects appear, the causes of the differences will be ascribed to heredity. Conversely, when heredity is held constant, and differences in effects appear, the causes of the differences will be ascribed to environment. However, the fundamental truth which is always to be kept in mind, and of which both the preceding statements are but corollaries, is that each and every outcome of the life process is at each instant of time the joint product of heredity and environment which are thus but two aspects of a single unity, life.

Heredity may be said to endow the individual with a specialized mechanism whose distinctive characteristics are its tendencies to act and grow in definite ways. These tendencies are potentialities, not actualities. Experience is the general term for the process by which potentialities are converted into actualities. When the same experience affects two individuals differently, the cause of the difference must be sought in the individuals and not in the experience. The potential power to respond to experience thus needs a name. Capacity will be the term employed. That is, genetically, capacity is that which is fixed by heredity; descriptively, capacity is potential development.

The concept, capacity, is given various names by different writers. Thus, Thorndike says:³

"What a man is and does throughout life is a result of whatever constitution he has at the start and of all the forces that act upon it before and after birth. . . . His 'original nature' is thus a name for the nature of the combined germ cells from which he springs and his environment is a name for the rest of the universe, so far as it may, directly or indirectly, influence him. In one sense nothing in human nature is due exclusively to either one of these factors. . . . But in another sense the most fundamental question for human education asks precisely that we assign separate shares in the causation of human behavior to man's original nature on the one hand and his environment or nurture on the other. In this sense . . . we **take for granted** the cooperating action of **one** of these two divisions **in order to think** more successfully and conveniently of the action of the other."

Cooley states the matter in very similar terms:⁴

" . . . we get capacity from heredity, conduct from society. While it is true that no amount or kind of education will take the place of initial capacity, it is true also that there is no source of right development and function except social teaching. . . . Practically, the two factors are always found in cooperation, and our knowledge that they are separable is largely derived from the lower forms of life where the social process is absent. . . . No doubt what we get in the germ cell is ours for life, and environment can only control or perhaps suppress its development. But it is equally true that heredity cannot overcome environment."

The statements of Sir Francis Galton made many years ago express the same fundamental distinction.⁵

"The interaction of nature and circumstances is very close, and it is impossible to separate them with precision. Nurture acts before birth, during every stage of embryonic and pre-embryonic existence, causing the potential faculties at birth to be in some degree the effect of nurture. We need not, however, be hypercritical about distinctions. We know that the bulk of the respective provinces of nature and nurture are totally different, although the frontier between them may be uncertain, and we are perfectly justified in attempting to appraise their relative importance."

Therefore, the fundamental causes of differences in individuals affected by like experiences will, in this study, be ascribed to

³Thorndike, E. L. *Educational Psychology* (1913) Vol. 1, page 2, 3. The italics are mine.

⁴Cooley, C. H. *Social Process* (1922) pages 203-6.

⁵Galton, Sir Francis. *Inquiries into Human Faculties* (1883) Page 131.

differences in inborn, original nature or capacity, and capacity will be assumed to be fixed by heredity. (See Chapter VIII for a discussion of the measurement of capacity.)

Environmental Forces. Experience acting on capacity develops both the physical mechanism and its power to function.

Even the very simplest of the known forms of life consists apparently of a complex mechanism which reacts in a definite way to the play of energy upon it. In most forms of life, however, the genetic process results in an immediate physical mechanism, the fertilized egg cell, which, though complex, is much simpler than the mechanisms of the adult parent forms from which it was derived. Under such conditions, living results in a period of rapid physical expansion, or growth, until the adult form is completely reproduced. This process of development, however, differs in its physical characteristics so markedly from the genetic process that it is given a number of different names.

Training. Experiences, while essentially all of one character consisting of the interactions of the mechanism and the rest of the universe, may be divided for purposes of classification and description into two types: controlled and uncontrolled. Controlled experiences are those deliberately entered into for the purposes of producing changes in the mechanism or its functioning, as when a man changes his diet in order to reduce his weight, or practices a stroke in golf in order to improve his game. In this study, controlled experiences will be called training and will mean in most cases school training. (See Chapter VIII for a discussion of the measurement of training).

Maturity. All other experiences such as eating, sleeping, going to a movie, etc., will be called uncontrolled experiences, because ordinarily they will be entered into for other purposes than to secure the ultimate effects produced. Thus, a boy eats to satisfy his hunger, not often that he may grow. Yet eating automatically and inevitably results in the development of the mechanism.

Experiences acting on capacity gives rise to ability. The effects of uncontrolled experiences are usually taken for granted

so that the term ability is ordinarily thought of as if used to designate only the effect of controlled experiences. Nevertheless, both controlled and uncontrolled experiences co-operate in determining ability. The general effect of uncontrolled experiences, especially such effects as reveal themselves largely in apparent changes in the physical mechanism with age will be called **maturity**.

The concept of maturity differs with different writers. Thorndike says:⁶

"No competent student doubts that in certain mental traits maturity or inner mental growth causes one individual to differ year by year from his former self, irrespective of all training."

Whipple gives a different definition. He says:⁷

"By maturity I shall mean that stage of development, that degree of unfolding, of this constitutional intellectual endowment which has been attained by virtue of the ripening of the individual's general mental efficiency brought about by the mere passage of time. If we could conceive of a given child at two years of age who then proceeded to live for one year without the slightest contact with the environment, the alteration in his mentality that would have taken place by the end of that year would be the effect of the factor maturity operating alone."

The physical organism is seen to change year by year until it reaches the status in size and form of the adult. These changes take place inevitably as a result of the fundamental life processes; they appear to be beyond control. No one "by taking thought can add one cubit unto his stature." Moreover these changes are seen to affect the functioning of the mechanism in the abilities which are the direct result of school training. There is probably little intrinsic difference in nature between development caused by controlled experiences and those caused by uncontrolled experiences. Practically, however, the concept of maturity, the effect of natural development as distinct from the effects of training, the deliberate manipulation of experience, serves a useful purpose. (See Chapter VIII for a discussion of the measurement of maturity).

⁶Thorndike, E. L. *Educational Psychology*, (1913) Vol. III, page 270.

⁷Whipple, G. M. "Endowment, Maturity and Training as Factors in Intelligence Scores," *Scientific Monthly* (1924) 18:499, May.

Ability. Ability may be defined as general power to do. However, one can never do anything in general; one always acts in particular situations. Thus, a man has, let us say, ability to write. To give his ability expression he has to write at a given time, on certain materials with certain tools and under certain conditions of lighting, health, etc. His actual writing on any occasion is thus a product of the joint action of ability and conditions. Thus, any behavior is not a pure expression of ability but of ability modified or conditioned by circumstances.

Ability may be considered as two quite different things:

1. As an existing, effective, but abstract power to achieve apart from limiting conditions; and
2. As the existence of a mechanism capable of achieving.

Thus a child's ability to add may be thought of as the power to carry on the adding activity, but what this really means is that his organism has been so affected by his past experiences that when the proper stimuli are presented there will be the adding response. This second viewpoint emphasizes the existence of a mechanism in which more or less permanent adjustments have been made. As long as the child has this mechanism unimpaired he will have general ability to add. His performance in any specific situation will be the theoretical power of his mechanism to add, modified by the specific factors of the situation. The second viewpoint is the one adopted for this study.

Conditions. For analytical purposes, the general mechanism, which we know as an individual, will be considered as consisting of many types or classes of simpler, unitary mechanisms four of which merit discussion. Every specific act of behavior will be considered to involve the functioning of three mechanisms, which may be described vaguely as the motor, mental, and emotional mechanisms, each of which is considered unitary because it is capable of a certain degree of independent or specialized development.⁸ The essential co-ordination of these mechanisms is accomplished through a fourth or control mechanism, which adjusts the internal to external conditions. Each act

⁸Compare with Thorndike's classification, Intellectual (abstract ideas), Motor (skill with hands), and Social Intelligence (ability to get on well with one's fellows), *Harper's Monthly Magazine*, January (1920) page 228.

of behavior, from this point of view, is the resultant of an intricate co-ordination or interaction of four types of mechanisms to stimuli.

In detail these four mechanisms are as follows:

1. The gross bodily mechanism, composed of skin, flesh, bones, muscles, the sensory organs, the transmitting nerve fibers, etc., the physical movements and reactions of which in space and time constitute the objective aspects of performance.
2. The mental mechanism, or that portion of the central nervous system whose interactions constitute the psychical counterpart of motor reactions.
3. The emotional mechanism, or the autonomic nervous system whose reactions are the physical causes, or determinants, of feeling.
4. The control mechanism or that portion of the central nervous system involved in conscious recognition of meaning, deliberate adoption of purpose, and voluntary direction of actions essential to achievement of purpose.

From this point of view, for instance, the general ability of a particular child to write is determined by the degree of organized control he ordinarily possesses over (1) his muscular movements; (2) the thoughts and ideas about written record desired;⁹ (3) his feelings about writing or not writing in the given situation, and (4) the specific mechanism which exerts control over the other mechanism involved. His actual performance, when asked by his teacher in a classroom on a certain day to prepare a sample of his handwriting for exhibition purposes, will be expressed in a record which is the resultant of the influence of the external and internal conditions which always affect the functioning of the human mechanism as a whole. (For a discussion of the effect of conditions, see Chapter VII.)

Performance. The product of ability acting under a given set of conditions will be called performance. Performance is thus a specific achievement, or a unit portion of behavior. Performance is always a complex of ability and conditions and therefore a performance has little meaning apart from the conditions under which it is secured.

⁹As forms of letters to represent sounds, slant, thickness of line, spacing, etc.

The functioning of the human mechanism is conditioned by three types of factors:

1. External physical conditions as light, heat, nature of environment in which change is made, etc.
2. Involuntary internal states or conditions due to past or present experiences, as fatigue, discouragement, excitement, drunkenness, etc.
3. Voluntary internal states or conditions caused by adjustments and coordinations of control, as methods of work adopted, degree of effort put forth, motive actuating activity, etc.

Variation in any factor whatever will cause a performance to change from one time to another, but it is assumed that "ability" is a measureable entity apart from the factors enumerated above and that performance will fluctuate in correspondence with fluctuations in any or all of these elements, even when the ability, which is the determining factor in the performance, remains constant.

Ability cannot be measured directly but must always be inferred. That is, if conditions are constant, ability is proportional to performance. Ability might be defined in different ways, as, maximum performance or mean performance. In any ordinary series of performances the author considers median performance as the best measure of ability. It represents with the least error the most probable score on the text trial.¹⁰

Statistical Methods. Scores in tests are records of performance, but partial records only. Thus, if in handwriting we measure rate and quality only, we are omitting measurement of eye movements, of the amount of energy expended, breathing, etc., yet each of these is a measurable effect of the writing activity, and a child's score is incomplete without them. For practical purposes we ignore very many aspects of performance and

¹⁰While statistically the probable error of the median is greater than the probable error of the mean, this statement is true only of the records of performance, not of performance itself. The true probable error, could it be determined, would be a function of both the probable error as statistically determined and the relative frequency of occurrence of the causes of variation. Specifically, if in a given class, two children make lower scores than usual, one because he sneezed in the middle of the test and the other because his pencil broke, the mean weights these variations more than the median, so that the median will give the score on the next trial in which neither of these variations occur, more truly than the mean.

Use of the median tends to decrease the effect of extraneous variation. Once reliable scores have been obtained, however, the mean gives the more probable measure of central tendency. Accordingly the mean is used through this thesis except in situations in which it is desired to make selection from non-homogeneous data.

measure those that seem significant. Yet such omissions tend to limit the completeness and accuracy of both measurement and inference.

Records and tabulations of performance are also influenced by still another set of factors, namely, the methods used for scoring, tabulating analyzing, and interpreting data. These affect the results and conclusions drawn from them. They are, therefore, included in the list of factors. Study of the validity and reliability of all statistical procedures is to be considered an integral part of this investigation.

Verification of prediction by experience is the criterion adopted for judging the truth of conclusions reached.

Summary. The foregoing definitions and discussions are summarized in the following condensed statement in which each important factor considered is numbered.

1. The genetic aspects of the reproductive process results in a partially developed and potentially specialized mechanism. (Heredity) In other words, the fertilized egg cell has a capacity (1) for development and modification.
2. Experience acting on capacity results in the development and modification of the inherited mechanism. The effects of physiological developmental experiences and unorganized social experiences are called Maturity (2).
3. Experiences organized to produce development or modification of mechanism or ability are called Training (3).
4. The power of the inherited and modified mechanism to respond to stimuli is Ability (4).

There are four types of mechanism involved in any ability, each of which is modifiable somewhat independently of the others. These are:

- The control mechanism (4a),
- The emotional mechanism (4b),
- The mental mechanism (4c),
- The motor mechanism (4d).

5. The action of the mechanism complex is affected by the Conditions (5) under which it acts. These are:
 - External physical conditions (5a),
 - Internal states (5b),
 - Internal adjustments and coordinations (5c).

6. Ability acting under given conditions results in Performance (6).

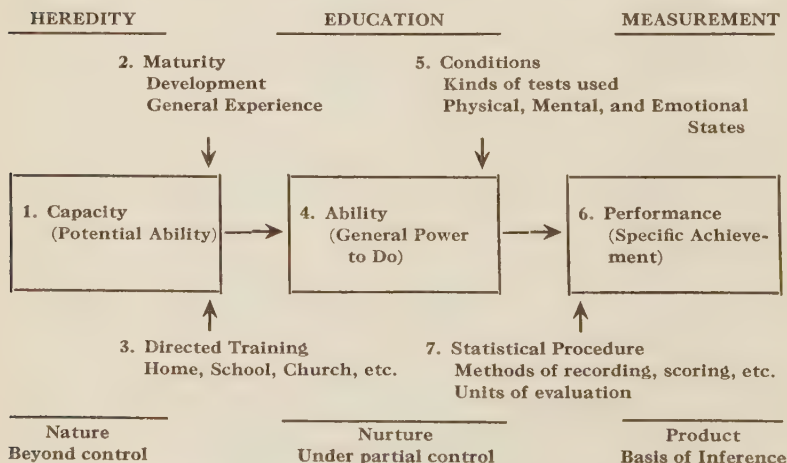
Performances are measured and recorded as scores. Any score, therefore, is merely an index of the resultant of the action of many factors of which ability is but one. Performance is a reliable index of ability only when the ability to be inferred has been precisely defined and allowance made for the effects of all other factors.

7. Inferences and generalizations are made from statistical records (7) of performance. They are affected, therefore, by
 The method of reducing performance to records (7a),
 The methods of summarization and generalization (7b).

The relation between the seven major types of factors affecting either success in school, or the measurement of success discussed in the foregoing paragraphs, is given in the following diagram.

FIGURE 2

Diagram of the Relation between the Seven Major types of Factors Contributing to Conclusions in Educational Measurement



Capacity is determined by heredity and is largely beyond control. Maturity and Training acting on capacity result in ability which is thus under partial control. The process of developing ability is Education. Ability functioning under specific conditions results in performance, and it is partial records of performance which, tabulated and analyzed, furnish the basis of inference.

The present study can deal only with the effects of gross outstanding factors; no attempt will be made to determine the effect of minor elements except to indicate how such analysis will ultimately be achieved. However the foregoing definitions and discussions should serve to make plain the sense in which the different terms are used.

CHAPTER V

The Measurement Activities and the Details of Procedure

The specific complex whose analysis is attempted in this dissertation is the progress, or success, of children in school as measured by the Stanford Achievement Tests and by a test of handwriting ability. To determine degree of success and the factors which contribute to it, the children in three public schools of Detroit were measured with these tests and with various other physical, mental, and educational tests. Records of teachers' marks, attendance, promotions, etc., were collected as well as a number of general and specific estimates of ability and emotional traits. For certain selected cases a study of home conditions was made. From time to time as need arose, special experiments were tried at nearby schools. Thus six types of data were secured: the results of physical, mental, educational, emotional, social, and administrative measurements.

Not all of the tests given had to do with the problems of this investigation but as data from many tests will be used freely as needed, it seems best to present the entire testing program at this point. Certain other investigations are still being continued. Only such of the data as are pertinent are reported in this study. The remainder is on file at Detroit Teachers College.

Program. The investigation was begun in November, 1922. Work has been carried on continuously from that time to the present (May, 1925). The program of activities has been as follows:

TABLE III
Program of Work

Month	Work Done	Pupils Concerned
November	Making of detailed plans and ordering of supplies.....	
December	{ Assembling of school records of various types and making of copies for use in this investigation; figuring of promotion rates.....	2,800
January		
February	Music Appreciation tests, form 1.....	2,000
	Teachers' estimates of capacity and of achievement.....	2,000
	Preliminary experiment, emotional survey.....	200
March	Experimental work in construction of new tests.....	500
April	Three measurements of rate and quality of writing.....	2,200
	Measurement of ability in oral reading (Gray) ages 10, 11 and 12...	700
May	Measurement of ability in silent reading	
	A. Simple rate test—(Detroit) Grades 4-8.....	2,000
	B. Simple Comprehension Test (Courtis) Grades 4-8.....	2,000
	C. Performance Scale (Thorndike-McCall) Grades 6-8.....	1,300
	Measurement of spelling ability, Grades 3-8.....	2,000
	A. Easy words	
	B. Difficult words	
	Measurement of adding ability	
	Four trials of special Detroit test, Grades 4-8.....	2,800
	Emotional Survey. Grades 2-8.....	2,000
	Measurement of Educational Age (Stanford Achievement Tests)	
	Grades 3-8.....	2,500
	Measurement of Intelligence	
	A. National Intelligence Tests, Grades 3-8.....	2,200
	B. Army Intelligence Tests (Detroit edition), Grades 4-8.....	2,000
June	Music Appreciation Tests, Form 2, Grades 3-8.....	2,200
	Sociological Investigations, Ages 10, 11, and 12.....	700
	Physical Measurements	
	Characteristics, height, weight, lung capacity, grip, Grades 2-8....	2,500
	Stunts—distance throw, target throw, running leap, twenty yard dash, Grades 2-8.....	2,500
	Motor—mental tests (Detroit special), Grades 4-8.....	2,000
	Composition tests, Grades 4-8.....	2,000
July, 1923-		
March, 1924	Scoring of tests, tabulation and analysis of results	
March 24	Measurement of Educational Age, Stanford Achievement Tests....	324
	Measurement of Intelligence	
	A. National Intelligence Tests.....	324
	B. Army Tests.....	324
	Motor-Mental Tests	
	(Seven selected classes for special experiment).....	324
May 26	Measurement of Intelligence	
	A. National Intelligence Tests, Grades 4B-7A inclusive.....	1,115
	B. Army Tests	
	B. Army Tests, Grades 4B-7A inclusive.....	1,115
June 2	Measurement of Educational Age	
	Stanford Achievement Tests. Grades 4B-8A.....	1,375
	Motor-Mental (Detroit), Grades 4B-8A.....	1,375
June 9	Physical Measurements	
	Height, Weight, Lung Capacity, Grip, Distance Throw, Target Throw, Running Leap, Grades 4B-8A.....	1,375
July, 1924	Measurement of Composition	
to	Ability, Rate and Quality, Grades 4B-8A.....	1,375
May, 1925	Measurement of Ability to Add	
	Four Tests, Grades 4B-8A.....	1,375
	Measurement of Handwriting	
	Ability, Grades 4B-8A.....	1,375
	Scoring, Tabulation and Analysis of Results	

All tests were given by carefully trained paid examiners and scored under supervision.¹ It was necessary for the purposes of the investigation to give the tests as nearly at one time as possible. On the other hand, interference with school work could not be carried too far. The many thousands of scores secured were recorded on 3" x 5" cards and used repeatedly in different types of tabulations. The figures in Table III above are round numbers only; absence and other causes operated to decrease the actual number of available records.

Conditions. The scope of the investigation was limited to children of ages 8 to 15 in grades 3 to 8 of the three training schools connected with Detroit Teachers College. The period covered is that in which the fundamental school arts and skills, commonly known as the three R's are acquired. During this period the general development curves for most abilities are quite rectilinear.

The children in the schools selected for the experiment are, in general, average or a little better in their social status, although as in many American schools, individuals may be found among them who come from "foreign" homes of poverty and ignorance, as well as from homes of culture, luxury, and the best American traditions. As a whole, however, the children represent middle class rather than exceptionally high or low social levels.

The distribution of children by age and grade in the May tests, 1923, is given in Table IV.

In each grade in each school children were constantly entering and leaving. At the close of the investigation, there were on file records from 3269 different children. Of these 1032 were enrolled at the time of the first testing but not at the second; 1323 were present at both periods; and 914 were enrolled at the time of the second testing period only. The actual available records complete for three or more different tests at both periods are very much less than the number in the table.

¹See Appendix for illustrations of certain tests and records and for details of giving and scoring.

TABLE IV
Age Grade Relationships Among Children Tested
Number of Children of Each Age and Grade for Whom Detailed Progress
Records Were Available

Age in Years	A1	B2	A2	B3	A3	B4	A4	B5	A5	B6	A6	B7	A7	B8	A8	Total
7.0	6	20	1	1												28
7.5	27	28	16	10												81
8.0	10	13	37	28	13	2										103
8.5		6	18	43	40	10										117
9.0	3	4	10	26	38	33	10									124
9.5				9	22	25	52	3	0	0	2					113
10.0			2	8	6	17	36	34	6	1	2					112
10.5			4	2	7	11	21	27	49	13	1	2				137
11.0		1	0	1	4	1	19	28	30	31	9	3	1			128
11.5				1	2	3	5	6	21	24	26	13	1	1		103
12.0		1	1	0	0	1	3	6	17	24	27	24	9	5	2	120
12.5						1	1	3	11	15	16	21	27	16	10	121
13.0				1	0	0	4	6	5	7	13	19	24	24	11	114
13.5						0	1	1	0	5	6	11	17	29	27	97
14.0						1	0	1		4	3	12	16	9	15	61
14.5							0	1		2	4	4	12	14	25	62
15.0							1			0	5	5	7	16	8	42
15.5											2	1	4	4	10	21
16.0											1	0	0	5	3	9
16.5													2	0	3	5
Total	46	73	89	130	132	105	153	116	139	126	117	115	120	123	114	1698

The records and measurements on file are as follows:

- | | |
|--|--|
| <p>I. School Records</p> <p>City Tests</p> <p>School Progress</p> <p>Teachers' Marks</p> <p>Teachers' Estimates</p> <p>II. Physical Measurements</p> <p>Height</p> <p>Weight</p> <p>Lung Capacity</p> <p>Grip</p> <p>Distance Throw</p> <p>Target Throw</p> <p>Running Leap</p> <p>Twenty Yard Dash</p> <p>Special Defects</p> <p>III. Mental Measurements</p> <p>Motor-Mental</p> | <p>Army Intelligence Tests</p> <p>National Intelligence Tests</p> <p>IV. Educational Measurements</p> <p>Stanford Achievement Tests</p> <p>Music Tests</p> <p>Handwriting Tests</p> <p>Adding Tests</p> <p>Reading Tests</p> <p>Gray Oral</p> <p>Detroit Rate</p> <p>Courtis Kitten</p> <p>Thorndike-McCall</p> <p>Spelling Tests</p> <p>V. Survey of Emotional Characteristics</p> <p>VI. Survey of Social Conditions</p> |
|--|--|

The reasons for the choice of these measurements are given in the appendix.

The methods of tabulation and analysis used were for the most part the conventional methods.³ Whenever departures from regular practice occur, they are noted, and explanations and illustrations provided in the appendix.

³Rugg, H. O. *Statistical Methods Applied to Education* (1917).
Kelley, T. L. *Statistical Method* (1923).

CHAPTER VI

Justification of the Adoption of a Narrow Definition of Success

Basis of Definition. The first step in the determination of the factors which condition success is the measurement of success itself, and the first step in measurement is the definition of the quantity to be measured. Success in school work may be defined from many different angles: from the point of view of society or of the individual, from that of the teacher or that of the home, and so on almost without end.

All definitions in terms of ultimate effect, however, involve the consideration and adoption of a specific philosophy of life. For instance, before one can determine whether a given change in a child will contribute to his success as a citizen, one must decide what success as a citizen means. This in turn is dependent upon one's definition of the purpose of life. The goal of living is stated differently by different schools of philosophy, and even if a decision were arbitrarily made by me in favor of a particular philosophy, it would be quite impossible at present with our limited skill to measure the effect of success in school in terms of the defined ultimate success in life.

Definition. Success, for the purposes of this investigation, must be used with a narrow connotation. It will be taken to mean the achievement of the immediate aims of the school in terms of development of ability in the fundamental school arts—reading, writing, arithmetic, etc.

Justification. However, this restriction of meaning is not as serious as it may at first sight seem. For when the schoolmen know what factors must be controlled to achieve **any imme-**

mediate goal, and what the relations between the factors are, they can apply that knowledge to the achievement of other goals if other goals prove more desirable. Even if the new goals should involve the control of one, or many, new factors, the methods of analysis, and the general conclusions from a study of present-day success, ought to be largely transferable to the new situation. In this study, therefore, success is defined from the point of view of achievement of the immediate goals of the educational process itself, rather than from that of the ultimate effect of such success upon either society or the individual.

Measurement of Success. The child who attends school is not wholly a free agent. He is "sent to school" by his parents. Schools are social institutions created by the collective action of parents for the good of their children. It might be expected that the report the school returns to the parents would be the best criterion of success to adopt. Parents, however, are easily satisfied, and the unreliability of teachers' marks has been shown by many investigators.^{1,2} That they are rough measures of general success, and as such are entitled to some consideration, has been proved by other investigators.^{3,4} But all the studies agree in showing that school marks are influenced by many subjective factors whose effects would tend to obscure the specific influence of a given objective factor.

Evidently what is needed is a measure of success which is narrow, objective, and readily controlled. Educational tests provide such measures. Accordingly, it was decided to define success in terms of score in educational tests.

Tests Chosen.⁵ A canvass of available tests led to the selection of two as suitable for the purposes of this study, the Stanford Achievement test and a test of handwriting. The first is a battery of tests covering reading, arithmetic, science, literature, language, and spelling. The complete total score is thus a measure of general school success in terms of absolute achievement.

¹Starch, D. *Educational Measurement* (1916) pages 3-15.

²Monroe, W. *Measuring the Results of Teaching* (1918) pages 1-21.

³Dearborn, W. F. "The Relative Standing of Pupils in the High School and the University." *Bulletin No. 312*, University of Wisconsin (1909).

⁴Kelley, T. L. *Educational Guidance* (1914).

⁵See Appendix.

The test was relatively new and so was unknown to the children to be tested. It had been devised and standardized by persons⁶ whose training and reputation were a guarantee that the tests themselves would represent as reliable measuring instruments as our present knowledge of test and scale construction permit. An additional advantage was that standards for comparison with other schools and children were available.

The test of handwriting was chosen for several reasons which are the exact opposites of those determining the choice of the Stanford tests. Handwriting itself is a relative simple and homogeneous ability. Its measurement involves the use of a rate test,⁷ whereas the Stanford tests are all performance scales.⁷ The use of these contrasted forms of measurement provides opportunity to determine which form of testing instrument yields results most suitable for analytical purposes. Also, other standard tests were available by means of which it was possible to secure the additional information needed to evaluate or interpret the results of these two basic tests.

Objections. The use of scores in standard tests as measures of success is open to one serious objection; they do not correspond to any of the conventional measures of success. That is, "Success," as defined for this investigation has a special and technical meaning. The child who makes the highest score becomes the most successful, even though he is counted a failure by his teacher or his parents. Similarly, the child who makes the lowest score is to be taken as the least successful, no matter how he may be rated in school.

Age, grade, intelligence, home conditions, and effort, as such, should have no more bearing upon the **measurement** of success than they have upon the measurement of height. The tallest boy in the school is the tallest on an absolute fact basis. Similarly, the most successful boy in the school, under the definition adopted, is the most successful on an absolute fact basis. The purpose of the study is to discover the part age, effort, and other

⁶T. L. Kelley, G. M. Ruch, and L. M. Terman.

⁷Hines, H. C. *A Guide to Educational Measurement* (1923) page 124.

factors play in determining such success. But they must play no part in conditioning its measurement. The degree of success attained at any specified time is a simple objective fact. Accordingly, the criterion of success adopted rests on a basis of absolute objective measurement so far as any measurement is either absolute or objective. This distinction should be kept in mind throughout the following discussions.

Justification of Measure Adopted. The decision to use scores in standard tests was not made arbitrarily, or without due consideration. It was made reluctantly, and only when it was found that no other basis would serve the purposes of the experiment. Therefore, it may not be out of place to present the evidence upon which the decision was based.

Difficulties in Using Teacher Judgment. If we select a single class, as the A5th in one of the schools, and find the mean mark given each child on the final school report at the end of the semester, January, 1923, we obtain, for the 37 children for whom complete records are available, the results shown in Table V.

TABLE V

Determination of Success in Terms of Teacher Judgment

Average (Mean) Mark* of 37 A5th Children in the Marr School in All Subjects Reported on Final School Records at End of Semester, January 1923.

Indi- vidual	Mark	Indi- vidual	Mark	Indi- vidual	Mark	Indi- vidual	Mark
1	1.4	11	1.8	21	2.3	31	2.8
2	1.5	12	1.8	22	2.4	32	2.8
3	1.5	13	1.9	23	2.5	33	2.9
4	1.5	14	1.9	24	2.5	34	3.0
5	1.6	15	2.0	25	2.5	35	3.0
6	1.7	16	2.0	26	2.5	36	3.2
7	1.7	17	2.0	27	2.5	37	3.4
8	1.8	18	2.2	28	2.5		
9	1.8	19	2.2	29	2.6		
10	1.8	20	2.2	30	2.6		

*School reports show marks of (1) Excellent, (2) Good, (3) Fair, (4) Poor, and (5) Failure, in the following subjects: Arithmetic, Drawing, English, Geography, History, Hygiene, Literature, Manual Training, Household Arts, Music, Nature Study, Penmanship, Helath Education, Reading, and Spelling.

That these marks are influenced by many subjective factors becomes clear if we ask one teacher to arrange the names of the children in the order of her estimate of their own achievement progress, or, in other words, in the order of her estimate of the degree to which the children had grown during the semester. (Table VI.)

TABLE VI

Comparison of Teacher's Estimate of Success with Average Mark

Table V with Ranks of Children on Basis of Teacher's Estimate of Success and with Marks Changed to Rank Order

Individual	Average Mark	Teacher's Estimate of Success	Individual	Average Mark	Teacher's Estimate of Success	Individual	Average Mark	Teacher's Estimate of Success
1	1	2	16	16	10	31	31.5	36
2	3	1	17	16	5	32	31.5	26
3	3	11	18	19	20	33	33	35
4	3	6	19	19	16	34	34.5	19
5	5	28	20	19	22	35	34.5	25
6	6.5	4	21	21	18	36	36	34
7	6.5	13	22	22	23	37	37	37
8	10	27	23	25.5	29			
9	10	17	24	25.5	30			
10	10	3	25	26.5	31			
11	10	9	26	25.5	21			
12	10	12	27	25.5	32			
13	13.5	7	28	25.5	24			
14	13.5	8	29	29.5	14			
15	16	15	30	29.5	33			

When a teacher observes that a child of great natural ability secures a relatively high achievement with little effort and with small gain, she rates him as less successful than a child of little natural ability who through great effort succeeds in making large gains, even though the final level of ability reached in the latter case is a relatively low achievement with reference to the class as a whole. The teacher's action seems justifiable to her because school marks are conventionally used as a means of reward and punishment, but to the degree to which they are so used, they cease to be pure measures of success. Any estimate

TABLE VII
Comparison of Various Factors with Average Marks

The data in this table consist of those of Table VI together with Differences in Sex, Age, Training, and Intelligence Tests. B means boy, G girl; +1 under age means a half year older than average; +1 under training means one half year more time spent in school than the average; under intelligence, figures are ranks corresponding to adjusted scores based on two group tests.

Individual	Rank Mark	Rank Success	Sex	Age	Training	Intelligence
1	1	2	B	-1	-1	5
2	3	1	G	+2	-1	10
3	3	11	G	0	0	3
4	3	6	G	0	0	1
5	5	28	G	0	-1	8
6	6.5	4	G	+1	0	8
7	6.5	13	G	0	0	24
8	10	27	G	+2	-1	14
9	10	17	G	0	0	20
10	10	3	G	0	-1	12
11	10	9	B	+1	0	6
12	10	12	B	+2	+2	8
13	13.5	7	G	0	0	3
14	13.5	8	G	0	0	18
15	16	15	G	0	0	18
16	16	10	G	0	0	36.5
17	16	5	B	+1	0	11
18	19	20	G	0	0	24
19	19	16	G	0	0	18
20	19	22	B	0	0	15.5
21	21	18	G	+4	0	27.5
22	22	23	B	0	0	3
23	25.5	29	G	+2	0	13
24	25.5	30	G	0	0	30
25	25.5	31	G	+2	0	36.5
26	25.5	21	G	0	0	34
27	25.5	32	G	0	0	15.5
28	25.5	24	B	-1	0	21.5
29	29.5	14	B	0	0	24
30	29.5	33	B	0	0	29
31	31.5	36	G	0	0	35
32	31.5	26	B	0	0	31
33	33	35	B	+3	+2	21.5
34	34.5	19	B	+2	+2	32
35	34.5	25	B	+2	+1	26
36	36	34	B	+1	0	33
37	37	37	B	+4	+4	27.5

Percentage in upper half of class—Boys 27%; Girls 66%.

of success which allows for the conditions under which the success is attained involves the making of many judgments of this sort, and also the arbitrary weighting of these judgments in arriving at a final estimate. Even if each element involved could be objectively measured, the mere problem of giving to each the proper weight would be an exceedingly difficult one. If we turn from subjective to objective factors, we find many other puzzling questions presenting themselves for consideration. The children differ in sex, age, and intelligence, and have attained the grade in which they are by various amounts of training. The average age for the class is 10.5 years and the average number of semesters of training is 12 (including two in the kindergarten). The ages range from 10 to 12.5 years. Five children have less than twelve semesters of training and five more. The intelligence scores used are adjusted scores based upon measurement with two group tests: the facts are given in Table VII.

The girls appear to do better than the boys, since $14\frac{1}{2}$ out of the 22 girls (66%) are marked in the upper half of the class, while only 27% of the 15 boys are in the upper half. If we take the teacher's estimate of success based upon growth in achievement, the figures become, Boys, 37%; Girls, 59%, which is a slight improvement but too small to remove the impression that the girls do much better than the boys. The teacher's estimate, however, agrees closely with the intelligence scores; for according to these, 40% of the boys and 57% of the girls should be in the upper half of the class.

The differences in age and training raise interesting questions. The boy and girl in first and second place according to school marks are of different ages but the same training. That is, the girl is more mature than the boy. Greater maturity may account for the reversal of order in the teacher's estimates, for the intelligence tests put the boy 5th, and the girl 10th in the class. The table makes clear the type of considerations which influence judgment of success, if the attempt is made to "make allowances." For instance, should girl four, who has achieved this position a year younger than girl two, with only one semester

more training and with the highest intelligence score in the class, be counted the more successful?

How complex are the age-grade-progress relationships is shown by consolidating the data for the various age groups (552 cases) in such a fashion that the facts of acceleration and retardation in terms of grade status can be analyzed to tell whether or not these are caused by early-or-late school entrance, or by slow-or-rapid progress (Table VIII).

Out of 552 children in grades 2 to 8, only 81 boys and 114 girls, or 35% of the total, are in grade at age. All the rest depart in some way from normal. Those within the heavy lines, or 55%, entered at the standard age and have made either rapid, normal, or slow progress according to their present status, but the rest of the children entered either earlier or later than normal and have made either normal or irregular progress. Here as elsewhere, there are sex differences. The girls come to school earlier and progress more rapidly than the boys. In evaluating the effect of training it may be that accurate account should be kept of the actual semesters each has been in school.

In Table VII, consider the five children in tenth place according to average mark. One of them (No. 12) is two semesters older and has had three semesters more training than another. A third one (No. 8), who is also two semesters older, has made the grade in one semester less than average time, yet is rated 27th by the teacher, while the first girl mentioned is rated 12th. According to the intelligence test, No. 12 is 8th in intelligence and No. 8, 14th. On the basis of teacher estimate, evidently some allowance for these differences would be made in determining success, but what allowance it is hard to say.

If to these data scores in the Stanford Achievement test are added (Table IX), interesting light is thrown on the situation.

Ten of the fifteen boys have higher ranks in terms of absolute achievement, as measured by the Stanford Achievement test, than in average mark, one has the same rank, and four lower ranks. In other words, it is apparent that in this class boys are in some way handicapped in receiving marks; that is, a boy

TABLE IX

Further Comparisons of Factors with Success as Determined by
Average Mark

Date from Table VII with rank in score in Stanford Achievement test added

Individual	Rank Mark	Rank Success*	Sex	Age	Training	Intelli- gence	Stan- ford
1	1	2	B	-1	-1	5	1
2	3	1	G	+2	-1	10	2
3	3	11	G	0	0	3	18
4	3	6	G	0	0	1	13.5
5	5	28	G	0	-1	8	3
6	6.5	4	G	+1	0	8	4
7	6.5	13	G	0	0	24	7
8	10	27	G	+2	-1	14	8
9	10	17	G	0	0	20	11
10	10	3	G	0	-1	12	10
11	10	9	B	+1	0	6	5
12	10	12	B	+2	+2	8	21
13	13.5	7	G	0	0	3	9
14	13.5	8	G	0	0	18	15
15	16	15	G	0	0	—	20
16	16	10	G	0	0	36.5	19
17	16	5	B	+1	0	11	16
18	19	20	G	0	0	24	12
19	19	16	G	0	0	18	24
20	19	22	B	0	0	15.5	33
21	21	18	G	+4	0	27.5	34
22	22	23	B	0	0	3	36
23	25.5	29	G	+2	0	13	13.5
24	25.5	30	G	0	0	30	25
25	25.5	31	G	+2	0	36.5	31
26	25.5	21	G	0	0	34	37
27	25.5	32	G	0	0	15.5	17
28	25.5	24	B	-1	0	21.5	22
29	29.5	14	B	0	0	24.5	30
30	29.5	33	B	0	0	29	28.5
31	31.5	36	G	0	0	35	36
32	31.5	26	B	0	0	31	28.5
33	33	35	B	+3	+2	21.5	35
34	34.5	19	B	+2	+2	32	27
35	34.5	25	B	+2	+1	36	26
36	36	34	B	+1	0	33	23
37	37	37	B	+4	+4	37.5	32

*Teacher's estimate of success.

must develop greater ability than a girl to receive the same mark.

That this is not a chance effect is shown by a tabulation of all the average marks given boys and girls (Table X). The general tendency is that girls as a class receive appreciably higher marks than boys.

TABLE X

Distribution of Average Marks of 819 Boys and 790 Girls,
Grades 2-8, on School Report Card at the End of the
Semester

	School Marks							Total	Median
	Low								
Mark.....	5.0	4.4	3.8	3.2	2.6	2.0	1.4		
Frequency									
Boys.....	2	9	15	194	331	205	63	819	2.86
Girls.....	0	1	5	86	277	324	97	790	1.95
Per Cent									
Boys.....	—	1	2	24	40	25	8	100	2.86
Girls.....	0	—	1	11	35	41	12	100	1.95

A still better way to show the same effect is to consolidate the distribution by ages of Stanford scores for each school mark in each grade, boys and girls separately (Table XI).

In general, in the schools tested, boys must have a higher absolute achievement to win the same mark as girls, the difference ranging on the average from 14% for a mark of 1, to 78% for a mark of 4.

These differences may be evaluated by comparing them with the differences between the absolute achievement which leads to the

TABLE XI

Comparison of Absolute Abilities as Measured by Scores in Tests
of Boys and Girls Who Receive the Same Average Mark

Mean score in Stanford Achievement test made by groups of girls receiving the average school marks indicated. Scores, differences, and percentage of difference are given to facilitate comparisons. Total number of cases 1609. A mark of 1.0 means an average from 1.0-1.499 of all school marks on the report card at end of semester. Values 1.5, 2.5, 3.5, 4.5 have been omitted from the table to save room but were computed. In finding percentages the scores made by girls have been used as bases. Results as given are correct to nearest whole number only.

TABLE XI, Continued

School Mark..	1.0 Stanford Scores				2.0 Stanford Scores			
	Boys	Girls	Differ- ence	%	Boys	Girls	Differ- ence	%
B8	110	100	10	10	85	72	13	18
A7	103	94	9	10	78	67	11	16
B7	97	87	10	11	73	62	11	18
A6	93	82	11	13	68	59	9	15
B6	88	77	11	14	64	55	9	16
A5	82	72	10	14	59	52	7	13
B5	76	66	10	15	53	48	5	10
A4	68	59	9	15	49	42	7	17
B4	60	51	9	18	43	37	6	16
A3	50	43	7	16	36	30	6	20
B3	41	35	6	17	29	24	5	21
A2	31	27	4	15	22	18	4	22
B2	22	21	1	5	16	14	2	14
A1	14	15	-1	-7	10	10	0	0
Midscore % of Difference				14	16.5			

School Mark..	3.0 Stanford Scores				4.0 Stanford Scores			
	Boys	Girls	Differ- ence	%	Boys	Girls	Differ- ence	%
B8	60	44	16	36	34	16	18	112
A7	55	40	15	38	30	13	17	131
B7	50	37	13	35	25	12	13	108
A6	45	35	10	29	21	11	10	91
B6	41	33	8	24	18	11	7	64
A5	38	31	7	22	16	11	5	45
B5	34	29	5	17	14	11	3	27
A4	30	26	4	15	12	10	2	20
B4	26	22	4	18	9	8	1	12
A3	22	17	5	29	8	5	3	60
B3	17	13	4	31	6	2	4	200
A2	13	9	4	44	5	2	3	150
B2	9	7	2	29	4	1	3	300
A1	6	6	0	0	2	2	0	0
Midscore % of Difference				29	77.5			

This table is read as follows: The mean score in the Stanford Achievement test made by the boys in the B8th grade who received an average mark of 1 (excellent), was 110; of the girls, 100. The difference in score was 10 points or 10% of the girls' score. On the average, for all grades, the score of the boys had to be 14% higher to receive a mark of 1, 16.5% higher for a mark of 2, 29% higher for a mark of 3, and 77.5% higher for a mark of 4.

Note that the ability for which a boy receives a mark of 1 in the A2nd grade (31), would earn him a mark of 2 in the B3rd (29), a mark of 3 in the A4th (30), and a mark of 4 in the A7th grade (30).

difference in marks. Thus, the A6th boys who received a mark of 1 made a score of 93, while the A6th boys who received a mark of 2 made a score of 68. But 93 is 137% of 68, a superiority of approximately 40%. The figures for A6th girls (82 equals 139% of 59) also show a 40% superiority.⁸ In other words, to receive the same mark from a teacher as a girl, a boy must be about one-third ($14/40$ or 35%) as much superior in achievement to the girl who receives the mark, as the achievement denoted by a mark of 1 is superior to the achievement denoted by a mark of 2. The average values for the 2—3, and the 3—4 differences figured on the same basis are 25% and 28% respectively.

Boys whose average mark in the upper grades is 4.0 (presumably "bad" boys) actually make scores in the Stanford test which are more than twice as large as the scores made by girls who receive the same mark.⁹

A further analysis of the data proves that for both boys and girls the relative ability to earn a given mark varies with age. To win a given mark, the brighter children in an age group (who are in a higher grade) must have a higher absolute achievement, while the older less able child (who is in a lower grade) is rewarded by the teacher with a higher mark than is given to normal children of that same ability. (Table XII, vertical columns.) In other words, the significance of a mark given to pupils of the same age changes from grade to grade.

The table is to be read as follows: The midscore Stanford score of four children 14 years old, one year accelerated, each of whom received an average mark of 2.5, was 73.5; of twelve children 14 years old, in grade at age, each of whom received an average mark of 2.5, 76.5, etc.

Note that twelve year olds, one year accelerated, (that is in a class with thirteen year olds), must have an ability represented by a score of 63.5 Stanford to win a mark of 2.5, while twelve year olds, three or more years retarded (that is, in a

⁸The mean value for all grades, boys and girls, is 40%.

⁹A significant article by Hughes, W. H. "Organized Personnel Research and Its Bearing on High School Problems," *Journal of Educational Research* (1924) X:386, December, makes it probable that these differences are caused by the differences in "traits and attitudes" between boys and girls, particularly in "respect for authority" and in "regularity, persistency, and trustworthiness."

TABLE XII
Relation of Maturity, School Mark, and Absolute Achievement in
Stanford Test for 178 Boys Receiving an Average Mark of 2.5

Grade Status	AGES							Total
	8	9	10	11	12	13	14	
2 years Accelerated						1		1
1 year Accelerated	1 6.8	1 28.4	8 31.3	1 66.9	8 63.5	5 62.3	4 73.5	28
In grade at age	2 14.5	18 20.6	17 36.4	12 46.6	11 54.0	16 65.0	12 76.5	88
1 year Retarded		8 22.6	1 16.4	4 41.6	9 54.6	9 58.8	7 71.1	38
2 years Retarded		2 16.7	1 23.6	3 35.2	5 54.8	5 60.6	4 65.4	20
3 or more years Retarded			1 20.3		2 56.5			3
Total	3	29	28	20	35	36	27	178

To compare the Stanford scores of children of different ages but in the same grade, follow the diagonals as shown by the arrows

class with nine year olds), need only have a score of 56.5 to win the same mark. If ten, eleven, twelve, and thirteen year old children are in the same grade,¹⁰ the ten year olds receive a mark of 2.5 for a Stanford 31.3, the eleven year olds receive the same mark for a Stanford of 46.6, the twelve year olds for a Stanford of 54.6, and the thirteen year olds for a Stanford of 60.6 ~~56.5~~. Although a teacher tends to accept lower scores from the younger children and higher scores from the older children in the same grade for a given mark, the differences they make do not offset the effects of differences in maturity so that the younger children receive less recognition than they should and the older more.

Note also that a mark of 2.5 given to eight year old children in grade at age has a Stanford equivalent of but 14.5, while the same mark for fourteen year old children in grade at age has a value of 76.5. In other words, the value of a given mark varies from grade to grade (horizontal rows), so that such marks cannot be used to determine laws without correction to a common standard.

The tabulation of Stanford scores in terms of school marks, grade, and progress reveals the average relative differences in absolute achievement in children necessary to cause school authorities to double promote, or to fail, them. Calling the Stanford score of children in grade at age 100%, the values of the scores of all other classes of children can be expressed as percentages of this standard (Table XIII).

The results obtained, however, would be a great surprise to anyone who was not familiar with the facts already presented. Some girls two semesters accelerated are, on the average, actually less able than those in grade at age, while girls one, two, three, or four semesters retarded, although less able than normal, do not show regular decreases in ability proportional to the semesters retarded. With boys the results are slightly more consistent, but only as much so as might have been expected from the fact that boys must be more able than girls to win the same recognition.

¹⁰See Chapter V, Table IV, page 55.

TABLE XIII

Relation of Acceleration and Retardation to School Achievement as Measured by the Stanford Achievement Test

Midscore of groups in grade at age is taken as 100%. In each grade the midscore of all other groups is expressed in percentages of the score of those in grade at age. Upper figures in each division, number of cases; lower figures, score. Based on records of 469 girls. Mean values only given for 451 boys.

Grade	Semesters Retarded					In Grade at Age	Semesters Accelerated	
	-4	-3	-2	-1			+1	+2
8		1 84	3 88	13 90		36 100	21 101	4 108
7	1 100	2 78	6 80	16 88		24 100	20 98	2 94
6	1 90	1 78	7 74	11 84		31 100	31 100	3 98
5			2 83	19 112		55 100	12 100	
4			1 82	9 101		39 100	17 104	
3			2 55	7 84		42 100	12 96	
2				1 38		11 100	6 53	
Total Cases	2	4	21	76		238	119	9
Mean Value	95	80	81	93		100	100	97
Boys	0	11	32	84		218	98	8
Mean Value	0	84	84	95		100	104	107

The explanation of these seemingly impossible results is, of course, to be sought in the facts already given. The accelerated children are for the most part young for the grade, and the retarded children old. Whether more able (or less able) than their own age group, they become relatively less different from the class average when promoted (or demoted) to other age groups. Further, retardation may also be due to bad behavior, absence, late entrance, etc., and acceleration to the opposite causes. Such changes are made without reference to the absolute abilities of the children. As a result, children of very different abilities in terms of absolute achievement are found in the same grade, or to put the statement in a more significant form, children of very nearly any grade level of measured ability from low to high may be found in every grade after the first four.

Effect of Objective Criterion. The adoption of scores in the Stanford tests as measures of success converts the complex situation which has been presented into a relatively simple one in which it is easy to see that the problem is to determine the reasons for success by evaluating the effect of the various factors. For instance, the data of Table VI, reorganized on the basis of rank in Stanford scores, together with additional data, suggest that the factors acting are very many (Table XIV).

The data show clearly several tendencies. Emotional control¹¹ is evidently a factor influencing success in this class. In the half of the class having the highest scores, there are but three children who are rated in the lower half of class in emotional control; individuals 5, 22, and 8. In terms of the objective measure of success adopted, these three children rank 3rd, 6th, and 8th from the top. In intelligence, they are 8th, 3rd, and 14th. In their school marks their ranks are 5th, 22nd, and 10th respectively, and in the teacher's estimate, their ranks are 28th, 23rd, and 27th. In other words, both school marks and the teacher's estimate of success are, in these cases, lower than the actual ability of the children would justify, because the teacher's judgment is warped by the misbehavior of the children.

It is interesting to note that, in general, those who rank

¹¹For method of rating emotion control, see Appendix.

TABLE XIV

Relation Between Objective Measures of Success and Various Factors

Ranks of Previous Tables Rearranged in Order of Success (Score in Stanford Achievement test) with Additional Data: Attendance (Days) During Semester, Physical Development in Relation to Height and Age, and Rank in Order of Emotional Control

Individual	Rank Success	Rank Teacher's Estimate	Rank Average School Mark	Rank Intelligence	Rank Attendance	Rank Weight for Height	Rank Emotional Control 1 = Best	Sex	Difference from Average	
									Age	Training
1	1	2	1	5	17	18.5	6	B	-1	-1
2	2	1	3	10	9.5	28.5	2	G	+2	-1
5	3	28	5	8	25	31.5	23	G	0	-1
6	4	4	6.5	8	6.5	35	8	G	+1	0
11	5	9	10	6	28.5	31.5	3	B	+1	0
22	6	23	22	3	20	25.5	23	B	0	0
7	7	13	6.5	24	31	13.0	18	G	0	0
8	8	27	10	14	2	23.5	26	G	+2	-1
13	9	7	13.5	3	30	18.5	13.5	G	0	0
10	10	3	10	12	20	36	6	G	0	-1
9	11	17	10	20	9.5	7.5	18	G	0	0
18	12	20	19	24	28.5	18.5	18	G	0	0
4	13.5	6	3	1	13.5	28.5	1	G	0	0
23	13.5	29	25.5	13	35	28.5	13.5	G	+2	0
14	15	8	13.5	18	32	7.5	18	G	0	0
17	16	5	16	11	25	14	9	B	+1	0
27	17	32	25.5	15.5	34	18.5	18	G	0	0
18	18	11	3	3	27	37	4	G	0	0
16	19	10	16	36.5	6.5	1	6	G	0	0
15	20	15	16	18	22.5	25.5	10.5	G	0	0

TABLE XIV (Concluded)

Individual	Rank Success	Rank Teacher's Estimate	Rank Average School Mark	Rank Intelligence	Rank Attendance	Rank Weight for Height	Rank Emotional Control 1 = Best	Sex	Difference from Average	
									Age	Training
12	21	12	10	8	17	18.5	10.5	B	+2	+2
28	22	24	25.5	21.5	20	5	26	B	-1	0
36	23	34	36	33	33	3	26	B	+1	0
19	24	16	19	18	2	33	30	G	0	0
24	25	30	25.5	30	25	18.5	32	G	0	0
35	26	25	34.5	26	12	9	13.5	B	+2	+1
34	27	19	34.5	32	4.5	11.5	34	B	+2	+2
30	28.5	33	29.5	29	13.5	22	28.5	B	0	0
32	28.5	26	31.5	31	4.5	2	33	B	0	0
29	30	14	29.5	24	17	23.5	35	B	0	0
25	31	31	25.5	36.5	22.5	6	31	G	+2	0
37	32	37	37	27.5	9.5	34	37	B	+4	+4
20	33	22	19	15.5	9.5	10	13.5	B	0	0
21	34	18	21	27.5	36	28.5	23	G	+4	0
33	35	35	33	21.5	2	15	36	B	+3	+2
31	36	36	31.5	35	15	4	28.5	G	0	0
26	37	21	25.5	34	37	11.5	21	G	0	0
Correlations:										
Success	—	.59	.79	.73	-.09	-.40	.64	Values of r were computed by Spearman's rank order method.		
Marks	.79	.77	—	.74	-.06	-.42	.76			

highest in success are rated high in emotional control, and vice versa; also, that ten out of the nineteen children in the better half of the class are under weight for their height, while thirteen of the nineteen are in the lower half of the class in attendance.

For the records of this class, the correlation of school marks and score in the intelligence tests is of the same order of magnitude as the correlation of Stanford scores (success) and intelligence score. The advantage of the measure of success adopted over school marks lies in the greater certainty of meaning. Each score is a positive objective fact, determined by the conditions and the child with the influence of subjective factors related to the examiner reduced to a minimum. School marks, on the other hand, are the product of so many subjective factors in both teachers and children that it becomes exceedingly difficult to trace the effects of the particular factors it is desired to investigate.

Conclusion. No general conclusions are possible from the records of a single class. But similar records from many classes confirm these results and prove that subjective marks and estimates are so distorted by the influence of many extraneous factors as to be valueless for purposes of precise evaluation. Accordingly, for this study, the decision was made to define success in objective terms.

Interpretation. However, this decision does not mean at all either that other types of success are not important, or that the conclusions from this study are to be taken as applying necessarily to the attainment of such other types of success. It does mean simply that, owing to the limitations of our present powers, it has been found necessary to put upon the investigation restrictions as to the type of success which was measured. The purpose of the measurement remains unchanged, and the investigation of relationships between objective success, as thus defined, and various factors is to be regarded as paving the way for the future investigation which will deal with measurement of success on other levels.

CHAPTER VII

The Measurement of Performance

Point of Attack. The first point of attack upon the problem of determining "Why Children Succeed" in terms of objective measurement of factors must be, necessarily, upon the process which yields the qualitative and quantitative records of those factors. Inaccuracies and inadequacies of record affect the outcomes of **all** subsequent analyses.

The Nature of the Problem. The major factors which determine success—heredity, maturity, training, and ability—are not subject to direct observation or measurement by direct comparison, as are such quantities as length, color, etc. The same is true of the four types of mechanisms described in Chapter IV, and of many of the conditions which affect their functioning. The sole observable and directly measurable products of the functioning of vital processes are of the movements of objects in space (behavior), or the resultant effects of such movements.

For example, when a child learns to appreciate a poem, the growth of appreciation cannot be directly observed or measured. What we do see are the changes in the behavior of the child which are caused by the changes in his appreciation. What he says when he appreciates a poem, differs from what he said before he learned to appreciate it; now he chooses to read the poem, hangs it in his room, or learns it by heart, etc., in preference to carrying on other activities which then he cared more for than he did for the poem. Saying and doing may be directly observed and have measurable aspects. But neither saying nor doing in themselves constitute appreciation; they are activities of considerable intrinsic magnitude and stability which are

merely modified in certain particulars by the existence of appreciation, so that only a part, and perhaps but a small part, of the total saying and doing is indicative of appreciation. Furthermore, appreciation is not the only element which may operate to modify either saying or doing.

The general problem, therefore, is to measure and so analyze a unit of complex behavior that discovery may be made of distinctive portions which may be used as a basis of reliable inference in regard to the influence of causes whose action cannot be observed directly.

Up to this point, the problem in education does not differ in any particular from the corresponding problem in other sciences. Energy, for instance, is a concept. It, also, cannot be directly observed or measured. It is an abstraction, a generalization from the analyses of the movements of objects in space. The direct observations of primitive man yielded neither energy, nor any other of the unitary concepts with which modern scientists deal effectively with behavior in the physical world. The elements of the alchemists were fire, water, earth, and air, complexes all; and the story of the rise and fall of the phlogiston theory in regard to fire is an illuminating illustration of the method by which, through intelligent use of the factor concept, complex behavior in many fields has been brought under control.

The characteristic which apparently differentiates the problems confronting educational science from those of the physical sciences is that of dependent variation. In the physical sciences, factors may be varied independently for all practical purposes. In gases, for example, it is possible to hold volume constant, vary the mass, and determine the effect on pressure. Moreover, if the effect of pressure on volume of a gas, and the effect of temperature, have been determined separately, their combined effect proves to be expressible in some simple relationship owing to the independence of their natures.

In education, however, we are dealing with a vital process, and the characteristic of life is change. Life is a balance between disrupting and integrating forces, a balance which is being con-

stantly upset, and as constantly restored. In the last analysis, this statement is true of all manifestations of natural forces, but the process of change is much more evident in the biological world than in the physical. The human mechanism is stable only in the sense that a ripple on the surface of a stream is stable; a ripple persists but changes form at each fluctuation in the flowing water.

For instance, a child cannot undergo intellectual training without also becoming more mature physically, mentally, and emotionally. No learning takes place singly; the effects of each experience, as is well known, are multiple, and the development of each ability affects the functioning of all abilities, although sometimes only in obscure and roundabout ways that are difficult to trace. Experimentation under the law of the single variable is thus much more evidently impossible than in the fields of the physical sciences.

I have used the phrase "much more evidently impossible" advisedly. It expresses the well-known fact that even in the physical fields no two events are ever completely and precisely alike. The time factor always changes, and many other factors are known to change, but the sum total of these differences is negligible for ordinary human affairs. For instance, the weight of the earth changes each instant, owing to the steady fall of meteors from space; yet in hundreds of years, the increase in mass does not amount to enough to change the gravitational constant appreciably.

Apparent constancy in the quantities measured is the rule in the physical sciences; many times it requires special research to detect the variations. For instance, the length of an iron bar is the same from one day to another within such small limits that for practical purposes the variations are disregarded, and often are not even recognized. On the other hand, in educational science, apparent variation is the rule; special research is necessary to reveal the underlying constancy. Thus, the score made by a child in a writing test may vary greatly from one test to the next given in the same half hour. Basically, however, the problems are probably identical in nature, and the experience

of men in other fields suggests that, as seeming complexity of variation is resolved into the effects of factors whose influence and action are known, constancy and simple relationships will be as much in evidence in education as in other fields.

Fundamental Assumptions. The first step toward ultimate control is, however, the clear recognition that measurement, and analysis of educational quantities, never deal, and in the nature of things as we now know them never can deal, directly with the quantities themselves. Behavior is always an intricate complex. Measurement of the activity of adding, for instance, is measurement of a form of behavior to which have contributed heredity, maturity, physical, emotional, social, and a host of other factors or influences, of which past experience in adding is but one. Score in a particular standard test of adding is, therefore, only a measure of performance in that test; it is not a measure of adding ability, if ability is defined as in Chapter IV.

It is apparent, of course, that ability **may be** otherwise defined. For instance, in a recent publication, a mental function is defined as "an actually, or possibly, observable event in behavior. . . . The terms mental function, capacity, and ability will be used interchangeably."¹ From this point of view, each variation in performance is a different ability. Thus, to add short examples, or long ones, to add in school, or in a bank, to add mentally or aloud, are examples of a few of the many different kinds of specific adding abilities with which education must reckon. This view is held by many persons and the following quotations serve to indicate its general nature.

"It (attention) is, in each of us, not one but many, and always the result not of an abstract but of a series of concrete developments. . . . And so with all other mental and moral virtues, they are not general but obstinately particular."²

"Maturities, as we have appraised them, are not everywhere one and the same. Growth is specialized and has many phases. In so far as physical status, mental age, educational age, and estimated mental, physical, social, scholastic, and emotional maturities as here secured represent phases or types of growth, the conclusion

¹Hollingsworth, L. S. *Special Talents and Defects*, (1923) pages 4-5.

²Moore, E. C. *What is Education?* (1915) page 102.

apparent is that growth is not everywhere uniform and single, but varied and, in different degrees, independent."³

"A second error in opinion about mental relations . . . is the doctrine that some one function is shared by all intellectual traits, and that whatever resemblances or positive correlations the traits show are due to the presence in each of them of this function as a common factor. In so far as they have it they are identical. In so far as they lack it, they are totally disparate."⁴

The view opposed to that of Thorndike is the "general and specific" factor theory held by Spearman and others of the English school. Spearman's statement of his position is as follows:⁵

"This continued tendency to success of the same person throughout all variations of both form and subject matter appears only explicable by other factors lying deeper than the phenomena of consciousness. And thus there emerges the concept of a hypothetical general and purely quantitative factor underlying all cognitive performances of any kind. . . . The factor was taken, pending further information, to consist in something of the nature of an 'energy' or 'power' which serves in common the whole cortex. . . .

The brain may be regarded as able to switch the levels of its energy from one to any other group of neurones; the amount and the direction of the disposable energy regulate respectively the intensity and the quality of the ensuing mental process. . . . The terms of our psycho-physical hypothesis, there is need of ascertaining, not only the potential of the general 'energy,' but also the effectiveness of the specific 'engine' employed."

"In place of the two factors, . . . one general and the other specific, Thomson prefers to think of a number of factors at play in carrying out of any activity such as a mental test, these factors being a sample of all those which the individual has at his command."⁶

In quoting statements of extreme views held in the controversy which has raged for twenty years, there is always danger of misrepresentation. Gates, for instance, also says:

"These statements (specific maturities) should not be taken to mean that the growths are quite or even equally independent and unrelated. All are positively associated and at least we have ad-

³Gates, A. I. "The Nature and Educational Significance of Physical States, and of Mental, Physiological, Social, and Emotional Maturity." *Journal of Educational Psychology*: XV, 352, September (1924).

⁴Thorndike, E. L. *Educational Psychology* (1913) page 186.

⁵Spearman, C. *The Nature of Intelligence and the Principles of Cognition* (1923) pages 5, 346, 351.

⁶Brown, W. and Thomson, G. H. *The Essentials of Mental Measurement* (1921) page 188.

duced new evidence that between any two desirable traits, the correlation is positive."⁷

and Thorndike, who through his writings and his influence upon the work and thought of his students, has persistently directed attention to the specificity of abilities, writes:

"The notions of mental machinery which, being improved for one sort of data, held the improvement equally for all sorts; of magic powers which being trained by exercise of one sort to high efficiency, held that efficiency whatever they might be exercised upon; and of the mind as a reservoir for potential energy which could be filled by any one activity and drawn on for any other . . . have now disappeared from expert writings on psychology. A survey of experimental results is now needed perhaps as much to prevent the opposite superstition; for, apparently, some careless thinkers have rushed from the belief in a totally generalized training to the belief that training is totally specialized. . . . These experiments should at least put forever out of court the **two absurdities**—that improvement is in an absolutely general 'memory' and that improvement is absolutely restricted to exact and entire duplicates of the situation met in special practice."⁸

In a recent article he states:⁹

"The imperfections (in measuring instruments) may be of two sorts. First the responses measured by the instrument may not be representative of the whole intellect and nothing but intellect; the scores obtained may not give enough weight to certain factors or elements of intellect and may give weight to others which really deserve less or even zero weight. . . . In general practice, however, we rarely know the relation of any person's effort to his possible maximum effort."

These passages and others which might be quoted, seem to reflect a point of view much more nearly in accord with that adopted for this thesis than the former quotations would indicate.

Spearman, on the other hand, bases his theory in part upon a hierarchy of correlation coefficients which range from low to high, thus recognizing variation in the degrees to which the specific and general factors contribute to a given performance.¹⁰

The facts in the case are clear and agreed to by all; it is in the explanation of the facts that differences arise. Measurable forms

⁷Ibid, page 353.

⁸Thorndike, E. L. *Educational Psychology*. Vol. II (1913) pages 365, 397.

⁹Thorndike, E. L. "Measurement of Intelligence," *Psychological Review* (1924) 31:227-8.

¹⁰See also Woodrow's characterization of the various theories as multifocal, unifocal, and nonfocal. Woodrow, H., *Brightness and Dullness in Children* (1919), pages 220-3.

of behavior in specific situations under given conditions differ from similar forms of behavior in other situations or under changed conditions. For instance, adding examples spaced half an inch apart on the test sheet results in a score which differs from that obtained when the examples are spaced two inches apart. For the same group of persons the correlations between the two sets of scores, while high, is much lower than might have been expected. If one test is of adding and the second of subtracting, the correlations are positive but lower. If the tests are repeated a year later, the correlations are generally lower still.

The point of view adopted in explanations presented in this thesis conforms more to that of Spearman than to that of Thorndike, although it is identical with neither. The basic concept has already been stated: namely, that an adding test does not measure adding ability, but adding behavior, to which adding ability is but one of many contributors. When the spacing, or the conditions of adding are changed, the new situation calls out, as it were, a different sampling (to use Thomson's term) of the person's total abilities. The new complex which is measured by the score varies with each change in the number and intensity of the contributing factors. Thus, although adding ability remained quite constant, adding in a widely spaced test could result in a very different complex from adding in a narrowly spaced test, owing to the difference in the contributions made by the motor factors.

In other words, scores which measure performance in a test will be considered in this thesis to represent the contributions of many factors. The problem is to determine these factors, and the laws of their combination, in such a fashion that the behavior of the individual in future situations where conditions are known may be predicted.

The situation in education appears to resemble the situation during the transition period from alchemy to chemistry. At that time, in the study of combustion, much attention was being given to the substance burning and to the flames produced, two very evident concomitants of burning. Various theories were

advanced to account for flame. Then came the analysis of combustion in terms of mass, with the consequent discovery of oxygen, the development of a new element concept, and the determination of the law of combination. The final outcome was the ability to predict in very exact fashion the complete behavior in all cases of combustion. Now the young student of chemistry has his attention directed **away from** the flames of combustion as a significant basis of observation and measurement. For flames in combustion are completely explained by the concepts of matter and energy and their relationships.

So it may be in the present situation in education, and this thesis represents an attempt, not to prove a theory, but to build one. That is, the effort will be made to predict one situation from another by resolution of the first situation into unitary factors, not in terms of a general and specific factor theory but in terms of whatever persistent tendencies may be traced from one situation to the other. Hypotheses are necessary, and the basic one is that complex behavior is resolvable into factors. The position taken, therefore, rejects the specificity theory while admitting the proved fact that each new instance of behavior differs from all others.

The rejection of this theory has already been justified: its adoption closes the door to discovery. It is as if the early chemist had said, "Each type of substance is specific. Sugar differs from butter and each from water. Therefore, the concept of elements is untenable."

As a working hypothesis, the opposite view is tentatively adopted. "Each instance of behavior is specific. The vast array of specific behaviors, however, seem to have certain characteristics in common. A search will be made for tendencies which seem to operate in different situations, and the isolation of their effects will be attempted."

It should be particularly noted, however, that no assumptions are made as to what these elements may be. The position is taken that at present we do not know that there are elements, and it will be time enough to name the elements when they

have been discovered. Certain general concepts have already been dimly outlined. Heredity, maturity, training, ability, conditions are but terms which aid in description and classification. None of them as yet represents more than a potentiality.

Take ability to add, for example. Adding performance is specific. If performance is defined as ability, then there are as many kinds of adding ability as there are possibilities of changing the spacing of the examples on the test paper; for each such new test gives rise to a new specific performance. The opposite view is that each performance is a complex to which adding ability is one contributor.

The question at once arises, "How is adding ability to be defined?" Does it include vision? If a child who makes many mistakes in adding, is fitted with glasses and, on his next test in adding and on all subsequent tests, makes a much better score than formerly, is it to be said that he has improved in adding ability? On the other hand, if he is tested with two types of adding tests, then practices adding in one of the situations until he becomes very skilful, but when he is given a second test in the other situation, is proved to have gained very little, how is the lack of transfer to be explained? These and many similar troublesome questions demand solution.

No attempt will be made to define ability except in the general terms already given:¹¹ as a general power to do, or as a mechanism capable of doing. Ability, whatever it may ultimately prove to be, results from experience. If, then, a unitary tendency or power developed by experience can be isolated by analysis from many adding situations, and the nature and extent of its contribution to adding situations can be so determined that adding behavior can be successfully predicted, the unitary power so isolated will be called adding ability, and the unique organization of neurones whose functioning determine the unique characteristics of such adding ability will be called the adding mechanism. Reasoning by analogy, it is possible to say that if such a unitary adding factor ever is isolated, it will bear little resemblance to present concepts of adding ability.

¹¹See Chapter IV, page 46.

Such considerations together with the supporting evidence from direct experimentation give rise to

Problem 1—Validity

Nature of the Problem. Validity is defined as the degree to which a test "measures that which it purports to measure."¹² In the past, makers of measuring instruments have labelled their tests in terms of the activity carried on. A test which involved the adding activity was called an adding test, one which involved reading, a reading test. A number of reading tests soon appeared. Correlation of scores in one reading test with those in another yielded surprisingly low results.¹³

The conclusion was at once drawn that there are many types of reading abilities, each of which is specific.¹⁴ But it was also discovered that scores in tests which apparently involve chiefly one type of activity sometimes are determined by other activities.¹⁵ Hence the need arose to determine what a test measures by some method other than that of merely assigning a name to it. The result is the problem of validity.

The methods of solution adopted by those who hold the various theories in regard to the specificity of abilities are very different. Thorndike, for instance, says:¹⁶

"Call difficulty for paragraph reading a characteristic of a paragraph and questions about the paragraph, much of which produces a large percentage of wrong responses to the questions, and little of which produces few, the individuals concerned being the same.

"Call achievement in paragraph reading that thing much of which enables an individual to respond correctly to a paragraph and a question about the paragraph involving much difficulty for paragraph reading whereas an individual of less achievement could respond correctly only to a paragraph and a question about the paragraph of less difficulty."

In other words, Thorndike says in effect, "Call ability to do this specific stunt reading ability." He does not recognize the

¹²Otis, A. *Statistical Method in Educational Measurement* (1924) page 256.

¹³Richards, A. M. and Davidson, P. E. "Correlations of Single Measures of Some Representative Reading Tests," *School and Society* 4:375 September (1916).

¹⁴Courtis, S. A. *Measurement of Classroom Products, Reading* (1919) pages 263-4. Gates, A. "Study of Comprehension in Reading by Means of a Practice Experiment," *Journal of Educational Research* (1923) VII:39.

¹⁵Courtis, S. A. and Thorndike, E. L. "Correction Formulae for Addition Tests" *Teachers College Record*, XXI:3 January, (1923).

¹⁶Thorndike, E. L. "An Improved Scale for Measuring Ability in Reading," *Teachers College Record*, January (1916).

previous existence of reading ability in the world, and so has no need to prove that individuals selected on the basis of scores in his scale as being individuals of great ability in reading will be judged as exceptionally able readers by their friends and acquaintances. Nor would he be disturbed if scores in his scale proved to have a higher correlation with intelligence tests which involve no reading whatever than they did with other tests which involve reading activity.¹⁷ He arbitrarily defines reading ability, so that for him the problem of validity becomes merely the problem of determining the correlation of scores in other reading tests with scores in his scale.

Those who hold the factor theory find themselves at once in difficulties. To them reading ability is a power or mechanism, built up through reading experiences, and contributing to every performance which involves the reading activity. A good reader is a person who possesses a distinctive something which is not possessed by the poor reader, a something which is not only distinct from oratorical ability, let us say, but is the distinctive product of reading experiences. No test can be accepted by him as a reading test until it has been proved by correlation to measure reading ability as otherwise determined.

Difficulties. The difficulties that arise are of two types: (1) the ultimate need for arbitrary definition of the criteria of judgment of what reading ability is, in precisely the same fashion that Thorndike arbitrarily defined it; and (2) the difficulty due to complexity. Both these difficulties will be discussed.

The need for ultimate arbitrary definition is universal. All human knowledge arises by generalization from experience, and all terms are but symbols arbitrarily selected to stand for such generalizations from experience. The alchemists who defined water as an element, differed not a whit from the chemists who define oxygen and hydrogen as elements, and water as a compound. There is no difference in the procedure; the chemists have merely hit upon a more serviceable definition. Fortunately, the chemists were able to isolate both water and oxygen, and

¹⁷As a matter of fact correlations of Thorndike-McCall reading scores with intelligence scores are high, generally higher than the intercorrelations between a series of different reading tests, as the previous references have shown.

determine the relationship of both to other situations. They selected finally that definition which possessed the highest degree of predictive power with the simplest explanations. Unfortunately, in education, no such absolute isolation of elementary phenomena is possible as yet. Thorndike, therefore, has chosen to capitalize present certainty in preference to possible future benefit. Those who hold the opposite theory deliberately put off the day of arbitrary definition in the hope that some consistent and persistent tendency may be traced through many situations and ultimately accepted as the basis of definition.

Types of Validity Determinations. Once this definition has been adopted and labelled, no test may be accepted as a valid test of defined ability until it has been proved by correlation to correspond to it. Thus, this thesis may be described as the report of a search for objective criteria that will serve as bases for definitions. Once these have been established, the problem of validity involves merely a process of comparative identification. Until that time, the process of solution is one of selective identification. For example, the problem of finding a valid measure of reading ability is one of isolation of a persistent tendency which for reading corresponds to the theoretical description adopted in Chapter IV.

Justification of Theory Adopted. The point has been made that no test measures what has been defined as pure ability; every test measures only an instance of complex behavior to which pure ability is regarded as contributory. In view of the acknowledged impossibility of dealing with pure ability in isolation, it might appear foolish to attempt, even a procedure at variance with the simple, practical expedient adopted by Thorndike. Yet the possibility of analytical treatment has already been established experimentally,¹⁸ and the results appear sufficiently promising to warrant an extension of the investigation. To aid in making the point of view clear, one specific illustration will be presented of a type of situation which occurs again

¹⁸Courtis, S. A. *Measurement of Classroom Products* (1919). *Factors Affecting Performance*, Chapter VII, page 335. "Correction Formulae for Addition Tests," *Teachers College Record*, XXI:3, January (1920).

and again, and which suggests very strongly the possibility of the resolution of various complexes into elements.

I refer to the increase in relative variability¹⁹ and decrease in correlation with apparent increase in complexity. I shall use for data the age-height, age-lung capacity, age-dash, age-target throw relationships for the boys measured in this investigation.

The full correlation table for the age-height relationship is given (Table XV). The mean in both age and weight has been located, and an ellipse drawn on the table to indicate the contour line of the correlation surface whose extremes are one standard deviation from the mean. The correlation is, $r = .87$. If all that was known about a boy was the fact that he was a boy between the ages of 7.5 and 15.5, the best prediction that could be made of his height would be the average height of such boys, or 55.3 inches. The standard error of such predictions would be for these six hundred boys 5.34 inches.

However, stature is known to be a relatively unitary and stable quantity directly dependent upon maturity. The degree of correlation between age and height, shown in the table, makes it possible to say, that if height were predicted from age, the standard error of prediction would be reduced almost 50%, or from 5.34 inches to 2.58 inches. If a prediction were made on the basis of both age and an hereditary factor analogous to build, the standard error would be reduced to a very small quantity indeed, a quantity representing the play of minor forces which may be designated as idiosyncrasies in growth cycles, health, etc.

In other words, the table represents a relatively simple situation in which departure from perfect correlation between age and height is brought about by the action of one other major factor and a few smaller ones.

Data are available from this same group of boys (except for a few absences at the time of the measurements) for the other physical items mentioned—lung capacity, twenty-yard dash, and target throw. Lung capacity is, of course, an elementary

¹⁹If scales are referred to true zero points and the factors do not operate against each other.

TABLE XV
Relation of Height to Age: 661 Boys
Ages by half years

Inch.	8	9	10	11	12	13	14	15	Total
74								1	1
72									
70						1		1	3
68							1	1	2
66						1	2	6	17
64						2	4	4	20
62				2	1	3	6	4	34
60			1	1	4	8	11	2	47
58			2	5	9	16	5	1	69
56		2	6	10	20	7	4	1	83
54		2	7	14	7	9			95
52	2	11	16	10	5	3	1		93
50	10	19	8	13					90
48	18	19	6	2					62
46	11	10	5						30
44	10	2	1						14
42	1								1
Total	52	37	59	49	39	50	32	16	661

8 means from
8 years, 0 months,
to but not including
8 years, 6 months

42 means from
42.0 inches to but
not including 44.0 inches

Mean Age: 11.1 years

Mean Height: 55.3 inches

Standard Deviation:
Age 2.21 years
Height 5.34 inches

Correlation:
 $r = .87 \pm .007$

Position of Mean
shown by $\bar{x}$

Contour line,
ellipse for $Z = 1$ S. D.

and unitary item much like height, determined largely by height and age, but the measurement of lung capacity makes demands upon voluntary effort and intelligence. One's record depends in part upon how well one knows the trick of blowing into a spirometer and upon the degree of effort put forth. Perform-

TABLE XVI

Relation Between Age and Scores in Various Physical Tests,
Boys, Ages 7.5 to 15.5 Years

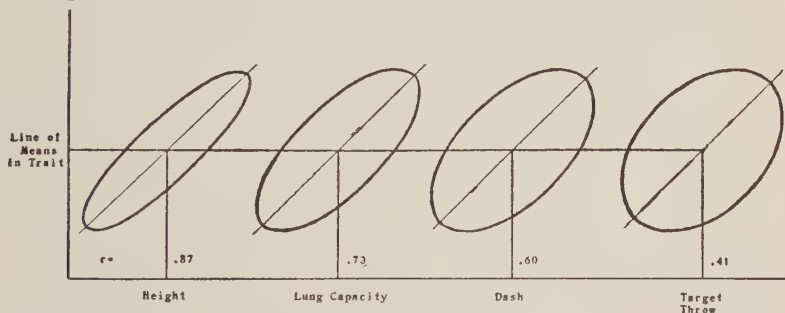
Test	Number Cases	Mean	S. D.	Relative Variability	Correlation with Age†
1. Height.....	661	55.3	5.34	9.7	.87±.01
2. Lung Capacity.....	659	123.4	43.8	35.6	.73±.01
3. Dash.....	654	4.03	.56	13.8*	.60±.02
4. Target Throw.....	629	14.2	6.64	46.6	.41±.02
5. Age.....	...	11.1	2.21	20.0	

†Pearson Coefficient.

*The score in the twenty-yard dash is a time score. The mean decreases in value with age, approaching a limit of approximately 2.4 seconds. The increase in variability is masked by the fact that the change in the mean is in a different direction than in the others and is not from the same kind of a zero point. If the minimum time is taken as the zero point the variability becomes 37.3%. The value in the table is therefore, not comparable with the others.

FIGURE 3

Comparison of Relative Complexity by Means of Correlation Surfaces



Diagrams of equal contour lines for correlation surfaces of performances of various degrees of complexity. The complexity increases from left to right. The increase in complexity is probably caused by an increase in the number of contributing factors.

The size of plotting units for standard deviations have been adjusted to make figures comparable. Ellipses were computed by formula:

$$\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} = \text{constant}^{20}$$

²⁰Brown and Thomson. *Essentials of Mental Measurement* (1921) page 124.

ance in a lung-capacity test is, therefore, more of a complex than performance in a height test, and the results show a greater variability and a lower correlation with age. (Table XVI, Figure 3.)

In similar fashion, the score in a test running twenty yards as fast as one can is even more complex; while the score representing the distance at which one is just not able to throw a base ball with precision enough to hit a basket ball (the target throw test) is more complex still. Each increase in complexity, the introduction of each new factor, increases the variability and lowers the raw correlation between the performance score and age.²¹ From the point of view adopted for this thesis, however, the change in correlation does not in itself indicate that maturity in height is different from maturity in target throw.²² The same physical maturity functions in all these performances, and the fact that it is not so evident in one performance as another, may be due to the disturbing effects of the other factors. What it does mean is that a lung-capacity score in a spirometer test is not a measure of lung-capacity ability as ability is defined in this thesis.

Conclusion. The conclusion drawn by me from these and other pertinent data is that the situation in education promises to be susceptible of analytical treatment of precisely the same genral type as that used in other fields, although modifications may need to be made in the analytical procedure to fit the special conditions; that the attempt needs to be made to determine predictive relationships and laws between persistent tendencies in performance, letting the naming activity follow and not precede isolation. The problem of validity is thus a basic problem.

Utility. The considerations which give rise to the problem of validity inevitably lead to the formulation of a second problem of especial significance for all practical use of measurement in school work.

²¹Brown, Wm. and Thomson, G. H. *The Essentials of Mental Measurement* (1921) page 99.

²²"The modern theory of correlation is directed toward the manipulation of observations made upon phenomena of a much greater degree of variability than that found in the case of isolated physical phenomena. The increased variability is no doubt due, in the main, to complexity of factors involved."

²²As Gates might say.

Problem 2—Utility

Nature of Problem. Performance has been defined as the product of ability acting under specific conditions. Performance in each test situation has many aspects. In general, man has tended to adopt the common sense procedure; he has selected and measured the elements of a situation which seemed most likely to be indicative of the ability he desired to measure.

Recognition of performance as a complex, however, leads to the conviction that not only must the search be for persistent and consistent tendencies which may be utilized as elements, but the relative worth of the different elements must be appraised. The problem of appraising worth is the problem of utility.

For instance, the measure of success adopted for this investigation is score in the Stanford Achievement test. For the most part, these tests call for the mechanical recall of items of knowledge, or the routine use of automatic skills. Conceivably, machines could be constructed that, in response to proper stimulation, could spell words, compute answers to numerical problems, or make the sounds indicated by printed symbols (reading). Knowledge has long been regarded as synonymous with power, but knowledge is frequently confused with ability to say appropriate words in response to questions. However, there are in the world elements of power whose contributions to performance are characterized by such terms as initiative, originality, executive ability, idealism, etc. If, now, elements contributing to mere word-knowledge performances were selected as the valid factors, the greater the degree of control over these factors that was developed, the more the product would approach pure word knowledge. Originality, initiative, etc., would tend to be completely eliminated.

In the physical sciences, the problem exists in no such degree. The chemist undertaking to create synthetically a natural product has the isolated product to use as a standard of judgment of success. The biological sciences, however, meet the problem in a simple form. When a plant breeder selects on the basis of

one quality that seems desirable, he sometimes finds that the selection involves undesirable changes in other qualities as well, so that there has to be a weighing of relative values. Nevertheless, these products of his activities are objectively observable and may be appraised in isolation. In education, however, the only products ever objectively observable are complexes, and special effort must be made to detect small traces of elements, which nevertheless are of supreme significance. Of course, in the long run, the product of educational activities is also susceptible of direct appraisal. A school whose graduates, year after year, become distinguished men would soon attract attention to the methods by which such success was produced. But the time between teaching and the test of successful living is so long, and the factors operating so many, that this means of measuring the value of school procedure, though valid, is difficult to apply.

On this general point, Thorndike says:²³

"Finally, it must be remembered that a very small spread of training may be of very great educational value if it extended over a wide enough field. If one hundred hours of training in being scientific about chemistry produced only one hundredth as much improvement in being scientific about all sorts of facts, it would yet be a very remunerative educational force. If a gain of fifty per cent in justice toward classmates in school affairs increased the general equitable use of a boy's behavior only one-tenth of one per cent, this disciplinary effect would still be worth more than the specific habit."

The problem of utility emphasizes the danger that lies in man's control of human affairs. If he plans wisely and succeeds, all is well; but if he plans poorly, the greater his success, the greater the ultimate disaster. Appraisal to determine utility should be, therefore, a constant feature of all selective activities, particularly of measurement activities, whose results are used as bases of judgment.

Appraisal of Worth. The standard of success adopted for this investigation is of unknown worth. While the problem is clearly recognized, means of solution have yet to be invented.

²³Thorndike, E. L. *Educational Psychology* (1913) II:421.

The outcomes and conclusions of the present study are likely to have, therefore, a narrow and questionable value. An effort was made to include a study of the effect of emotional and social elements, but the results secured are too inadequate for a true appraisal. However, the intent of the investigation as a whole is rather to isolate problems and develop methods of solution rather than to solve the difficulties which present themselves. Many years of research will be required to surmount even the obstacles brought to light by this simple investigation. On the other hand, it may well be that methods found effective for the isolation and measurement of certain abilities will be effective, also, in the study of the more complex performances. In any event, the problem of utility means, for the present study, caution in extending its conclusion beyond the narrow range of factors and products with which it deals.

Reliability. The problem of measurement of the products of school work may be approached from a third point of attack.

Problem 3—Reliability

Nature of Problem. In all human measurements there are at least three sources of error; errors due to imperfections in the measuring instruments themselves, errors caused by faulty use of the measuring instruments, and errors arising from the operation of relatively small chance or uncontrolled factors. Accordingly, any measurement, repeated, is likely to yield a different result. The amount of difference is determined by many causes. The correlation of the scores from two trials of a test, yields a measure of reliability.²⁴

The variations in score resulting from repeated testing are in some tests small. For instance, in one study height was measured to the nearest inch by school teachers after a brief training period, and remeasured three days later. In 85% of the cases, there was no variation, in 14% one inch of variation, and only in 1% of the cases did the variation exceed an inch. In weight, 50% of the cases did not vary more than one-quarter of a pound.

²⁴McCall, W. A. *How to Measure in Education* (1922) page 307.

Ruch, G. M. *The Improvement of the Written Examination* (1924) page 28.

Kelley, T. L. *Statistical Method* (1923) page 200.

But variations were more likely to occur in weight than in height measurements.²⁵ A study of the variations made in other physical measurements (lung capacity, grip, running leap, distance throw, target throw) shows that for one hundred cases tested two days apart, the mean error ranges from 5 to 10%, but the range of individual variation within each test is from 0% to 40% or 50%.

Explanations. The conventional method of looking at these errors is to consider them indications of imperfections in the tests.

On the factor hypothesis, however, another explanation is possible. Height may be measured very reliably, because it is the expression of relatively few factors which do not vary greatly. Weight is measured less reliably only in the sense that it is less constant. For instance, if a boy today weighs 96.25 lbs. and tomorrow 98.75 lbs., each weight may be determined with great exactness and reliability. The fluctuation in score of 2.50 lbs. is not an indication of unreliability in the sense of error of measurement, but only in the sense that performance in weight is subject to the play of factors which are not constant in their operation. This distinction is a most important one for educational measurement. In many situations, the products of performance are measured with great reliability; it is performance itself which varies.

The term reliability has been borrowed from other fields and given a new meaning in education. If thought of as "coefficients of constancy of performance," reliability correlations based on self-agreement would have a distinctive value, because they would indicate the presence or absence of modifying factors. For instance, if a child in one test writes at the rate of 110 letters per minute, a result which may be reliably determined to the last letter, and with a quality of 65 Ayres, a result which with care may be determined within one per cent, and tomorrow writes at a rate of 95 with a quality of 80, both of which scores are reliably determined, it is a mistake to call the handwriting

²⁵Elliott, C. M. "Reliability of Height and Weight Measurements." *Detroit Educational Bulletin*, 5:35, June (1921).

test unreliable. In this case, the sum of rate and quality ($110 + 65 = 175$), in the first test, exactly equals the sum of rate and quality in the second test ($95 + 80 = 175$). The test scores express the facts in regard to performance very reliably in each trial. The relation between rate and quality, however, changes, and the fact of change is an indication of a change in the influence of some factor.

It is important that tests measure reliably, but in such cases as the one just quoted, the test is not at fault. Thus, in general, there is need to distinguish between two types of unreliability: (1) that which is an index of inaccuracy of measurement of a definite quantity, and (2) that which is an index of the extent of the change in the factors contributing to the complex being measured. In the last analysis, all unreliability is of the second type, but the distinction is made to call attention to the fact that when variation in score occurs as a result of changes in the complex, such changes are significant, but when the variations are caused by extraneous, small chance factors which have nothing to do with the quantity being measured, they have a very different meaning and may be corrected by averaging.

Illustrations. The magnitude of variations in score from the true value is an important matter, as it affects subsequent analysis and conclusions.²⁶ The conventional way of increasing reliability is to lengthen the test by increasing the number of test items.²⁷ A better way is to determine the cause of variation and to control the factors which produce it. The first method raises the statistical measure of reliability but does not change the situation; the second modifies the situation so that the quantity measured is more consistently the same quantity in all measurements. Improvement in consistency of performance is proof of effective control of the factors operating.

As an illustration of the need for solution of the many problems involved in control of constancy of performance, a comparison of the scores of two groups of children in the Stanford Achievement tests will be given. With one group, made up of

²⁶Otis, A. *Statistical Method in Educational Procedure* (1925) pages 234-5.

²⁷McCall, W. A. *How to Measure in Education* (1922) page 310.

TABLE XVII

Comparison of Variations in Immediate and Delayed Retesting

Comparison of the changes in score in the Stanford Achievement Test made in two trials two days apart with changes made in two trials a year apart. Changes expressed as per cents of individual scores. + = gain in second test, — = loss. The second group was matched with the first in number, age, sex, score in the initial test and grade as nearly as possible.

Number of cases, Girls 36, Boys 42.

A. Equality of Two Groups in Age

Age

Group	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15 & Over	Median
GIRLS												
Two Days	2	5	4	10	6	2	3	1	2		1	11.8
One Year		6	5	10	6	1	4	1	2		1	11.8
BOYS												
Two Days	3	3	7	7	7	4	6	1	3		1	12.1
One Year	3	2	7	8	7	5	5	1	3		1	12.1

B. Equality of Two Groups in Grade

Grade

Group	5B	5A	6B	6A	7B	7A	Median
GIRLS							
Two Days			5	31			6A
One Year	1	7	8	15	3	2	6A
BOYS							
Two Days			5	37			6A
One Year	2	9	11	15	4	1	6B

C. Comparison of Changes in Stanford Scores

Per Cent of Change

Group	-12 or Over	-8	-4	0	4	8	12	16	20	24	28	32 or Over	Mean	S. D.
GIRLS														
Two Days		5	6	3	9	8	4	1					4.9	6.44
One Year			1	1	1	4	8	12	6	2		1	16.3	6.98
BOYS														
Two Days	1	2	12	9	7	8	1		2				3.7	7.14
One Year			1	1	3	4	11	8	6	3	4	1	16.8	7.86

Boys and Girls together, Means 4.25, 16.6

two sixth-grade classes in two city schools, the Stanford test was repeated two days after the first trial. Then from the regular tests given in this investigation, a group of scores were selected exactly matching the first in number, sex, and value, and as nearly as possible in grade. (Table XVII.)

The average gain during the year, expressed as a percentage of the initial score, was for the girls 3.3 times, and for the boys 4.6 times, as large as the change in score in two days, similarly expressed. Yet the standard deviations of the gains for the year are approximately the same as the standard deviations of the changes caused by repetition.

The explanation for the variations in score in the repeated testing is simple. Since the interval, two days, was too short to result in any real gain or loss in abilities that have been developed through six years of school training, the variations are clearly caused by variations in the contributions made by the various factors acting. In other words, the performance measured is inconstant to such an extent that any individual score, boys and girls together, is on the average likely to be 4.25% higher on the second trial. The standard deviation of the percentage gains and losses of the individual scores is 6.79%.

The explanation of the changes in score at the end of a year's training is less simple. Part of the variation in score may be due to the inconstancy of performance and part may be due to the variations in the effect of a year's training upon ability. The variability for the year is slightly larger than for the retest, but, on the other hand, it is possible that the practice effect of the first trial is negligible when the second test is given after an interval of a year.

Conclusions. The two certain facts are (1) that the variability of gains from a year's training is of the same order as the variability of the changes in a retest after an interval of two days, and (2) that the second variability may be the product of at least two variables, practice effect and growth, so that no positive conclusions may be drawn in regard to either. The most that can be said is that in attempting to predict individual success, as measured by the Stanford test from relationships

derived from this investigation, the departure of predicted from actual score due to uncontrolled factors **may** amount on the average to as much as 4%, and that the standard deviations of such errors of estimate may be as much as 6.79%. When to this amount is added the possibilities of similar variation in the other tests used, and when it is remembered that inconstancies of scores of this type lower correlations and interfere with the determination of relationships, it should be evident that the solution of the problems involved in rendering performance in a test constant are also fundamental to the solution of the general problem "Why Children Succeed."

Quality. Closely related to the foregoing problems, but one of a slightly different nature is

Problem 4—Measurement of Quality

Bases of Records. If the problem of test and scale construction is considered with the idea in mind of controlling variation in score through control of the various factors which produce variation, it will be found that behavior has three evident objective magnitudes which may be used as a basis of record: quantity, quality, and complexity, or, as it is more often termed, difficulty.

Measurement of Rate. The simplest measurement situation is when two individuals work at a uniform task and perform their work equally well. The difficulty factor and the quality factor are then constant, and difference in rate of performance, or amount done per unit of time, is the only record that need be considered in comparing the performances of the two individuals. Amount is the simplest of all quantities to measure. It is necessary only to adopt as a unit some division of amount—a word, or example, for instance—and by direct comparison, and a counting activity, to determine the total number of units accomplished in a given time. Time is easily measured. Instruments of precision are available for the purpose. The measurement of the simplest case of performance presents few difficulties.

Nature of the Quality Problem. Unfortunately, the simplest case almost never occurs. Even when the task is uniform,

the quality of performance varies, so that both quality and rate must be measured. Performance, defined as the movement of objects in space, or the objective effects of such movements, may range from mere uncontrolled, instinctive release of energy, as when a person laughs in response to a funny story, to the precise adjustment of matter to some ideal arrangement through exact control of the energy necessary to bring about the arrangement: as when a person moves a pencil over paper to make the marks which represent the sum of a series of numbers, or makes the sounds by which the ideas called forth by printed symbols are expressed. In the first cases, measurement consists in quantitative description of the response, how long the person laughed (duration), how hard he laughed (intensity), and so on. In all other cases, however, these characteristics are ignored for a more evident one, the degree of control exhibited in making the desired adjustment, or, as it is more conveniently described, the quality or accuracy of performance. For example, if the subject is attempting to find and write the sum of an example having an answer in four figures, and if he write two of the figures in correctly, his performance is of lower quality than that of an individual who has only one of the figures wrong.

In similar fashion, an individual who reads orally a printed paragraph, reads with a lower quality of performance if he mispronounces two words, than he who mispronounces one.

Immediately difficulties arise. How shall quality be measured? If a subject makes a single mistake in an example, has he not failed to do the task satisfactorily? If an individual reads a passage with vigor, appreciation, and expression, is not the quality of reading to be rated higher than that of a lifeless rendering of the same passage, even though the first reader make three mistakes in pronunciation and the second none? If the performance is an exceedingly complex one, in which many important characteristics may be distinguished, as vocabulary, syntax, organization, content, form, etc., etc., the difficulty of evaluating the quality of the performance either absolutely or relatively, is exceedingly great.

Solutions. The reaction of measurement men to this situation has been varied. For the most part, their behavior has

been determined by teleological reasoning. Rational human performance is purposeful; it is directed to the achievement of some end. Accordingly, almost without exception, standards for measuring quality have been developed on the basis of comparing the product as a whole with a product set up as ideal in terms of some purpose. For instance, the quality of handwriting, composition, drawing, sewing, etc., are measured in terms of general merit,²⁸ merit being defined in terms of the subjective judgment of the persons selected as competent judges. Merit, under such conditions, means "good" in terms of some ideal criterion, utility, beauty, conformity to a model, etc. It varies as the ideal varies. The quality of performance in adding, spelling, reading, etc., was in the early days of the measurement movement measured by the amount done perfectly, examples right, words spelled correctly, etc., being considered the only part of the performance which was of value. It should be evident that such criteria, while of value in estimating the utility of the product, are not adequate for judging the process by which the product is created.

The Ultimate Solution. No solution for the problem of measuring quality can be given in this thesis, except to point out that scientific thinking in this field has very largely missed the point. It has persistently studied the product in terms of the product, rather than in terms of the producer or the process. It seems probable that an ultimate solution will be attained only through securing detailed knowledge of three fundamental elements, the mechanism acting, the amount of energy released through the mechanism, and the degree of control exercised over the action of the mechanism. The discussions below will throw a little light on the situation, but not much. It only remains to point out that measurement of rate and quality is primarily measurement of "skill" of performance, and that the significant aspect of the situation is likely to be "degree of control." In other words, the problem will be solved when the

²⁸Thorndike, E. L. "Handwriting." *Teachers College Record*, 11:1, March (1910).

Hillegas, M. B. "A Scale for the Measurement of Quality in English Composition by Young People." *Teachers College Record*, 13:331, September (1912).

Thorndike, E. L. "The Measurement of Achievement in Drawing." *Teachers College Record*, 14:345, November (1913).

Murdoch, Katherine. "The Measurement of Certain Elements in Hand Sewing." *Teachers College Contributions to Education*, (1919) No. 103.

factors of amount of energy released, and the condition of the mechanism, can be held constant or allowed for, and quality measured wholly in terms of degree of control. In spite of evident defects, the nearest approach to such measurement of quality at present is the measurement of merit, or of accuracy of execution.

Difficulty. Consideration of quality of performance inevitably leads to consideration of the effect on quality caused by changes in difficulty. This gives rise to

Problem 5—Measurement of Difficulty

Nature of the Problem. The most frequent need in education is to measure the growth in ability of a student as he passes from one task to another, particularly in one field, as algebra or English, but also from field to field, as for instance, to compare his ability to write a composition in English with his ability to solve problems in algebra. Such measurements involve all the elements previously discussed with the added problem of evaluating the difficulty of different tasks in terms of some common scale.

Solutions. Not only is it not possible at the present time to offer any solution for the problem of difficulty, but it must be confessed that few of those who measure have recognized the problem as a problem. Even the concept of what is meant by difficulty has not been clearly formulated. Professor Thorndike has written what, so far as I know, is the only discussion of the problem, and excellent as the discussion is, it does little more than call attention to certain of the problems involved. He says:²⁹

“The new problems of theory and technique in the measurement of intellect, that is, the problems not soluble by the general methods of measurement in any science, concern the measurement of difficulty of task. Extent and speed are measurable in two of the most perfect units there are—number and time. In the case of difficulty, however, we have so far provided only for an inventory of intellectual tasks and their arrangement in the order of difficulty.

²⁹Thorndike, E. L. “The Measurement of Intelligence.” *Psychological Review*, 31:243, May (1924).

"Their difference in **amount** of difficulty and the difference between the amount of difficulty of any one of them and some zero point of difficulty (some task which is just below a task of infinitesimal difficulty), are not determined. To find ways of determining these will be our main work."

Present-Day Solutions. The two problems of measurement of quality and difficulty have been solved by inventing procedures, which, as Thorndike indicates, discriminate between degrees of quality and between degrees of difficulty, but do not measure either. These procedures will be discussed later. At this point, it may be well to point out that two types of tests have been devised for measurement purposes, the rate test for measurement of rate and quality, or briefly **skill**, but not difficulty, and the performance scale which measures difficulty but not skill. Each will be discussed in turn.³⁰

Rate Tests. Pure³¹ rate tests are instruments designed for the measurement of skill. They are made up of elements which conform to certain objective specifications, so that the elements are of uniform quality.³² If the subject is able to do one of the elements of such a test, he is able to do them all if given sufficient time. Rate tests are used, therefore, not to distinguish those who can do a certain task from those who cannot, but to determine the degree of control each individual possesses over an activity he is known to be able to carry on successfully.

The two aspects of behavior in a rate test that are usually considered significant are rate and quality: the number of elements an individual completes in a given time, and how well he completes them. There are many other aspects of performance in a rate test, but these are usually disregarded (1) partly because it is not known how to record them under ordinary school conditions, but (2) chiefly because, so far, none of these aspects have seemed to be closely related to ability. A few of them are pulse rate, breathing rate, effort put forth,

³⁰"Report of Standardization Committee," *Journal of Educational Research*, IV:78, June (1921).

³¹Pure used with a test is to be given the connotation "completely conforming to theory." Most tests depart from theoretical principles in some particulars.

³²The construction of rate tests, composed of elements of uniform difficulty has been explained in several places and accordingly the explanation is not repeated here: namely, Report of Committee of School Inquiry, Part II, Subdivision I, Section D, (1912), pages 16-19.

Measurement of Classroom Products (1919). Gary Survey, page 203.

Test and Scale Construction. *Scientific Monthly*, (1925). An article accepted, but not yet published.

emotional reaction to being tested, etc. The fact remains, nevertheless, that all of these are elements of behavior, and changes in them may be just as indicative of ability as the aspects that are usually considered. As new factors influencing scores in tests are discovered, it is probable that certain of these aspects may become significant.

Measurement by Rate Tests. Human behavior is a vital process in time. It doesn't stand still to be measured, but is always in process. Behavior produces change in material objects. Test makers seek for tasks in which some material product of activity remains after the activity has ceased. For instance, adding is an activity, the significant element of which cannot be isolated or observed except in action. Records can be made of the results of the adding activity, however, and it is these results that are really measured. For addition, the number of examples added can be counted and the correctness of the answers determined. A score of 16 examples tried and an accuracy of 50% constitute one method of describing the activity which produced them, and such scores are ordinarily accepted as measures of the ability behind the activity.

Rate tests are always timed, and the timing is adjusted so that the subject cannot finish the test. For if time enough were given, every subject who knew how to complete one element, or who was capable of learning how, would eventually complete the test successfully.

The situation is analogous to the use of trucks to move loads. If a one-ton truck and a five-ton truck are each used to move ten tons of brick a mile, each truck will eventually complete the task, the one-ton truck making ten trips and the five-ton truck but two trips. To compare the power (ability) of the trucks, however, they must be observed in transit, loaded, and under comparable conditions. If each is loaded to capacity, their powers are proportional to their loads. But if the one-ton truck is loaded to capacity because it is to run on a smooth concrete road, while the five-ton truck carries but one-fifth of a capacity load because it must haul that load over a country road of clay and mire, wrong inferences will be made as to the

relative power of the two trucks from the measurements of the loads alone, unless proper allowance is made for the difference in the conditions under which the two trucks are operating. So, too, with skill. If all other things are equal, the individual who can do the larger amount of work in a given time is the more skillful, or to put it in terms of previous definitions, has the better developed mechanism. The problem of measuring skill is fundamentally that of determining the degree to which the significant mechanism has been developed. Accordingly, test makers attempt to eliminate the condition factors by keeping them uniform, so that the behavior of the mechanism itself may be studied.

Types of Performance. The field of performance in relation to time may be divided into three regions: (1) the region of perfect control; (2) the region of variable control; (3) the region of no control.³³ Translated into the measurement of adding ability, these statements have the following meanings:

(1) **PERFECT CONTROL.** If sufficient time is allowed, those who can add correctly one example in the test are able to do every example correctly. They add and re-add as many times as is necessary. The region of perfect control extends from infinite time up to an interval so short that the individual does not have opportunity to check all his answers, and so has at least one answer wrong. No individual can perform an act like adding correctly unless sufficient time is allowed for the perfect functioning of the adding mechanism. For everyone the time interval can be made so short that the process is carried on imperfectly. When the time interval reaches the point for any one individual where his execution begins to be imperfect, the boundary between the region of perfect control and the region of variable control has been reached.

(2) **NO CONTROL.** If the time limit is made short enough, no one has time to add a single example. If an individual were hurrying to finish within a very short time limit, the functioning of the adding mechanism might be so imperfect that no single answer was correct. As the time interval is lengthened, a value

³³Courtis, S. A. and Thorndike, E. L. "Correction Formulae for Addition Tests," *Teachers College Record*, XXI:13-14, January (1920).

would eventually be reached, where at least one example of those tried would be right. This is the boundary between the region of no control and the region of variable control.

(3) **VARIABLE CONTROL.** The region of variable control is the region between the two boundaries just defined. In it a person's performance is more or less inaccurate, owing to the fact that the adding mechanism is functioning more or less imperfectly.

Measurement of skill is possible only in the region of variable control. That is, if two persons are measured using a time limit so great that each completes the entire test perfectly, the score of the two persons is the same. Yet one may have added each example five times and the other but twice. This aspect of their behaviors is not revealed by their scores. Consequently, the real difference in skill is concealed. But if the two persons are measured under a time limit, short enough to reveal the actual functioning of the mechanism but not too short to prevent it functioning altogether, the relative skills of the two subjects can be correctly inferred from their scores. If both have 75% of their answers right, and one does 20 examples while the other does 10, the one that completes the larger amount in a given time, other things being equal, is the more skillful. Or, if both complete 20 examples, but one has 50% of his problems right while the other has but 60% right, other things being equal, the one who works with the higher accuracy is the more skillful.

The difficult case to evaluate is when one works 20 examples and has 50% of them right, and the other solves 15 examples, a smaller amount, but has 75% right, a higher accuracy. Such cases, however, but serve to emphasize the fact that in the region of variable performance, behavior is two dimensional, like area. One cannot tell the area of a sheet of paper by measuring one dimension only. Neither can one arrive at a true measure of the degree of development of the adding mechanism by measuring either rate or accuracy alone.

Misconceptions. This is probably the place to comment on a common misconception. Some authors of textbooks on mea-

surement have advocated the use of the number of examples right as a measure of ability.³⁴ Others protest any form of measurement of rate or speed.³⁵

One of the reasons for this attitude has already been mentioned, appraisal of the product from the point of view of its utility. A wrong answer is valueless for any life purpose, and it is essential that all children learn that if they are responsible for finding the sum of a series of numbers, their responsibility is not discharged until the sum has been found *correctly*. The problem of measuring the skill of the person doing the adding is quite a different problem, and one which is not to be settled by arbitrary decisions founded on mere opinions, but by evidence. Therefore, the next step will be to submit the evidence on which the foregoing statements are based.

The points made, the existence of three regions, and the two dimensional nature of behavior in the region of variable performance, apply to all rate tests without exception. But rate tests in some subjects make evident certain of the items of importance, while rate tests in other subjects reveal other items. Addition tests are peculiarly useful to establish the idea of the three regions of performance, but because the examples form discrete units, study of variations in performance involves certain difficulties which will be discussed later. The field of behavior selected for specific study in this thesis is handwriting.

Proof of Variation. An individual tested under various conditions in his region of variable performance exhibits many different types of scores. (Table XVIII, Fig. 4.) As the subject was a mature individual who has been writing twenty-five years, it will be assumed that his actual ability in handwriting was not changed to any appreciable degrees by these tests. The ninth test, given after a lapse of eighteen months and falling between the results for tests four and five, proves that under similar conditions the individual performs in a consistent manner. These various scores, therefore, may be regarded as products of a single ability in handwriting, X, modified in different ways

³⁴Trabue, M. R. *Measuring Results in Education* (1924), page 285.

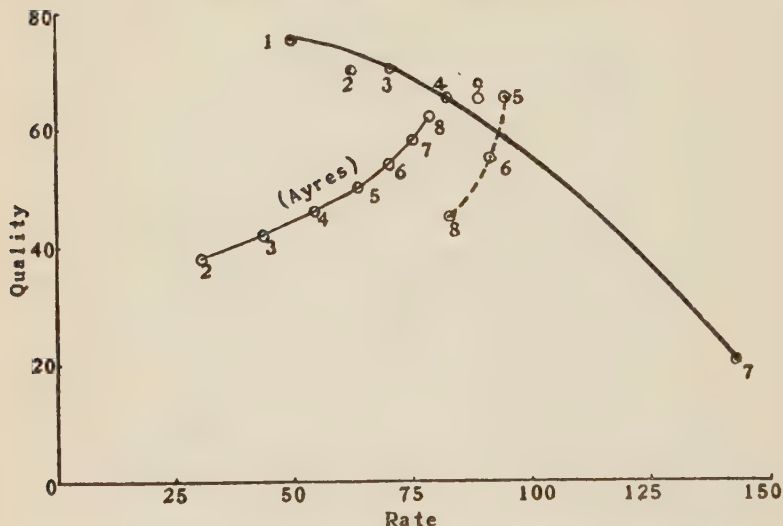
³⁵Unsigned editorial, "Speed and Accuracy," *Journal of Educational Research*, April (1925), page 292.

TABLE XVIII
Individual Variation in Performance
 Scores of One Individual in Handwriting Measured by Several Tests Under Various Conditions

No.	Test Material	Conditions	SCORES	
			Rate	Quality
1	Mary had a little lamb.....	Write as well as possible, disregarding rate.....	51	75
2	This is a sample of my best writing.....	Prepare a sample of best writing.....	63	70
3	Standard paragraph.....	Copy as well as possible, disregarding rate.....	72	70
4	Standard paragraph.....	Copy rapidly and well.....	84	65
5	Mary had a little lamb.....	Write rapidly and well.....	96	65
6	Lincoln's Gettysburg speech.....	Copy rapidly and well.....	92	55
7	Standard paragraph.....	Copy as fast as possible, disregarding quality.....	143	20
8	Standard paragraph.....	Copy rapidly and well, standing holding paper in hand.....	83	45
9	Standard paragraph.....	Copy rapidly and well.....	90	65

Note: Tests 1-8 were given within an hour. Test 9 was given eighteen months later.

FIGURE 4
Individual Variation in Performance



Scores of one individual in handwriting measured by several tests under various conditions

by the contributions of other factors, Y, Z, etc. The problem is (1) to discover the various factors acting, (2) to determine the laws governing their relationships, and (3) to derive from a series of different scores the single score that will most accurately represent the ability of the individual in handwriting in any situation in which he may be placed.

Testing Conditions. The children tested in this investigation were given three tests in handwriting. In the first, they were asked to write rapidly and well a printed test paragraph. This will be called the normal test (N), as it represents the testing conditions in general use. In the second test, they were urged to write as rapidly as possible (F, or "fast" test), paying no attention to quality of writing. The test was made a race to see who could write the largest number of letters in the time allowed. In the third test, the emphasis was put upon quality.

The children were urged to write as well as they could, paying no attention to the amount written (S, or "slow" test).³⁶

Conditions of Scoring and Tabulation. The papers were scored as to rate of writing by counting the letters per minute, and as to quality by expert scorers using the Ayres scale.³⁷

Great care was taken with the scoring, and the results are reliable within small limits of error. Thus, in testing the differences in quality scores assigned by independent scorers, the median of 905 differences was 1 point. (Table XIX.) In similar fashion, a recount of the number of letters written on some 95 papers gave results for which the median difference from the first scoring was 0 letters. In other words, the scores in the following tables are to be considered accurate within at least 3% on the average, and the variation discussed below is not to be ascribed to unreliability in the measuring instruments or to the method of reducing handwriting performance to quantitative measures. The scores truly reflect variation in the functioning of the writing mechanism.

Results. The scores of the children in the various tests varied greatly. Even when the records of children of the same sex, age, and grade are compared, they are found to vary widely.³⁸ In terms of the theory adopted for this thesis, these results can mean only that many factors contribute to each score.

Methods of Treatment. Similar variation in scores has been found by many investigators and various expedients have been adopted to reduce rate and quality scores to a single score. Thus, Thorndike, in addition, for each example wrong, deducts a certain amount from the number of examples right.³⁹ He gives no basis for the amount deducted, other than assumption. Brown, in reading, defines a unit of reading ability as follows:⁴⁰

"One unit of reading efficiency is a reading ability in which such rate of reading is combined with such a power of reproduction that

³⁶See Appendix, directions for giving handwriting tests.

³⁷See Appendix, procedure to secure uniformity in scoring.

³⁸See pages 123, 125.

³⁹Thorndike, E. L. *Educational Psychology* (1914) page 21.

⁴⁰Brown, H. A. "The Measurement of the Efficiency of Instruction in Reading," *Elementary School Teacher*, 14:482, June (1914).

TABLE XIX

Reliability of Scoring Rate and Quality in Handwriting

Rate scores are letters per minute; quality scores are in terms of the Ayers scale.

		Variation in Rate Scores											Total	
		1	2	3	4	5	6	7	8	9	10	11	12	Median
Amount of Change on Recount	0	1	2	3	4	5	6	7	8	9	10	11	12	
Frequency	49	32	7	3	1	1	1	0	0	0	0	0	1	95 0
		Variation in Quality Scores											Total	
		1	2	3	4	5	6	7	8	9	10	11	12	Median
Amount														
Variation in Score in Papers Accepted	-10	-8	-6	-4	-2	0*	+1	+3	+5	+7	+9	+11		
Frequency	2	6	14	38	271	225	275	54	12	5	2	1	905	1.2
		Variation in Quality Scores											Total	
		1	2	3	4	5	6	7	8	9	10	11	12	Median
Variation in Score in Papers Rejected and Rescored	-30	-25	-20	-15	-10	-5	+5	+10	+15	+20	+25	+30	+35	
Frequency	1	4	7	26	36	1	39	43	18	4	1	2	1	183 ..

*The interval is 2 in all cases except zero which here stands for perfect agreement.

†Without regard to sign.

Each paper was measured three times: once by a rank order method, the papers being in a large group, and twice by two judges independently. Differences of more than 10 points were supposed to be rejected and rescored by a third judge. On all other papers the mean of scores was taken as the true value. Total number of papers, 1088.

the product of the numbers representing rate and the average of the numbers representing the percentages of quantity and quality is unity."

However, no evidence is offered that this unit is related in any way to ability in reading, other than that it yields numerically a score, compounded from the various elements measured, which is a multiple of a defined unit score.

For cancellation tests, the methods used by various investigators are summarized by Whipple, who recommends, without justifying evidence, the formula:⁴¹

Efficiency index — Speed index (ground covered) multiplied by
accuracy index.

or Speed index (in terms of time) divided by accuracy index.

He also states, "Possibly an empirical corrective formula might be experimentally determined."

Starch has given more attention to the reduction of rate and quality scores to a single score than most other writers.⁴² He adopts a scheme of correction, uniform for different rate tests, as follows:

"If we may assume, as we shall do here, that a growth in speed from the first grade to the eighth grade is equal to the parallel growth in comprehension from the first to the eighth, the growth in speed (of reading) of 2.5 from 1.5 in the first grade to 4.0 in the eighth grade, is equivalent to a growth in comprehension of 35, from 15 in the first grade to 50 in the eighth. Hence a growth of .1 of a word in speed is equal to a growth of 1.4 words in comprehension."

In general, his plan is to convert the speed scores into comprehension, or the comprehension score into speed, calling the average of the two scores the real measure of ability. Thus, a score of Speed 3.0, Comprehension 25, would be changed to Speed 3.0, Speed 2.2, and the score for ability taken as 2.6.

In writing, a pupil whose speed of writing is 41 letters per minute and whose quality is 10 (Thorndike) would have a single score of 56 in terms of speed, since on the Starch conversion

⁴¹Whipple G. M. *Manual of Mental and Physical Tests* (1910) page 260.

⁴²Starch Daniel. *Educational Measurements* (1918) pages 33-34, 84.

scales 10 (Thorndike) is the equivalent of 70 in speed.

$$(70+41) \div 2 = 56.$$

The most recent attempt to reduce speed and quality to a single score is that of Gates.⁴³ The formula he proposes is of the general form:

$$\text{General Ability} = Q^t S^k \text{ where } t = 1 \text{ and } k = \frac{1}{3} \text{ or}$$

$$\text{Combined Score} = \text{Quality} \times \sqrt[3]{\text{Speed}}.$$

Gates says of this formula:

"The effect of the formula is to tremendously reduce the range (i. e., the weight) of speed. For example, for Grade V (Scarborough results) the average figures are:

	Writing for Quality	Writing Normally	Writing for Speed
Quality Score.....	10.375	9.41	8.75
Speed Score.....	45.7	64.7	73.3
$\sqrt[3]{\text{Speed}}$	3.57	4.01	4.23

"For widely differing rate scores, the formula transforms the results to small differences, thus greatly reducing the weight of speed on the combined score. To use the fourth root would practically eliminate speed."

How true the last sentence is may be seen by comparing the ratios of the quality alone with the ratios of the products of speed and quality, of $\sqrt[3]{\text{speed}}$ and quality, and of $\sqrt[4]{\text{speed}}$ and quality. The ratios of the values obtained by multiplying quality by the fourth root of speed approximate the ratios of the measures of quality alone.

	Speed	Quality	Ratio	Quality x Speed	Ratio
Speed Test.....	73.3	8.75	1.00	641.4	1.00
Normal Test.....	64.7	9.41	1.08	608.8	.95
Quality Test.....	45.7	10.375	1.19	474.1	.74

	$\sqrt[3]{\text{Speed}}$	Quality x $\sqrt[3]{\text{Speed}}$	Ratio	$\sqrt[4]{\text{Speed}}$	Quality x $\sqrt[4]{\text{Speed}}$	Ratio
Speed Test.....	4.18	36.7	1.00	2.32	25.6	1.00
Normal Test.....	4.01	37.8	1.03	2.84	26.7	1.04
Quality Test.....	3.57	37.1	1.01	2.60	27.0	1.05

⁴³Gates, A. I. "The Relation of Quality and Speed of Performance." *Journal of Educational Psychology*, XV:129, March (1924).

TABLE XX

Variation in Computed Scores

Comparison of Mean Scores of a Class in Eleven Handwriting Tests Given Within Ten Days by Four Different Examiners

Test	Ayres Rate-Quality	Thorndike Rate-Quality	Computed Score*		Per Cent of Variation Highest to Lowest	
			Courtis	Gates	Courtis	Gates
Teacher's First.....	61-57	61-11	118	44.0		
Student Teacher A						
Normal.....	63-59	63-11.5	122	45.3	3.39	2.96
Slow.....	51-64	51-12.3	115	45.2	6.9	
Fast.....	78-45	78- 9.6	123	41.5		8.9
Student Teacher B						
Normal.....	58-61	58-11.8	119	46.0	1.7	
Slow.....	55-65	55-12.3	120	46.0		20.4
Fast.....	78-43	78- 9.5	121	38.2		
Student Teacher C						
Normal.....	64-57	64-11.2	121	44.5		
Slow.....	53-59	53-11.2	112	43.0		
Fast.....	80-43	80- 9.2	123	39.2	9.8	13.5
Teacher's Second.....	71-59	71-11.4	130	46.9		
Mid-Points between slow and fast tests.						
Student Teacher A.....	64.5-54.5		119			
Student Teacher B.....	66.5-54		120.5			
Student Teacher C.....	66.5-51		117.5			
Teacher's First Test.....			118			
Teacher's Second Test.....			130			
Mean, All Tests.....			120.4	43.6	23.3	22.8

*Formulae for computed score: Courtis, Rate plus Quality (Ayres); Gates, $\sqrt{\text{Rate} \times \text{Quality}}$ (Thorndike).

In other words, the Gates index is a statistical device which yields the nearest approximation to a constant score when the true measure of writing ability is assumed to be a product of speed and quality. It throws no light on the cause of the variation in score, and no evidence is presented to validate the assumption on which it is based. Moreover, a tabulation of the scores made by 33 children in a series of eleven handwriting tests under various conditions proves that the Gates index does not yield a constant value. For instance, the indices for one child in the eleven tests computed by Gates' formula were: 32, 34.4, 28.0, 31.3, 34, 34, 28.1, 31, 26.5, 35, 36: mean value 31.8. The extreme deviation above the mean was 4.2, below 5.3, which is, using whole numbers, a range of 9 points. But 9 is 28% of 32. A similar measure of range was computed for each of the thirty-three children. The distribution and mid-value of the ranges, or per cent of variation, follows:

Per Cent Range Is of the Mean											
	10	15	20	25	30	35	40	45	50	55	60
Frequency...	1	0	2	8	7	6	3	3	1	0	1
Mid-Value 33											

Even the range of variation in the eleven class averages in the various tests was 23% (Table XX). On the basis of these data, Gates' formula must be judged inadequate even from the point of view of yielding a statistically constant index of ability.

What is true of the Gates' formula is true of the devices adopted by others. All of them are intelligent attempts to replace the awkward two-dimensional score by a single score, and all of them undoubtedly yield measures which approximate true measures of ability. None of them, however, is based on experimental analysis of the situation, only on more or less plausible *a priori* assumptions. None of them throws any light upon the causes of the variation they serve to eliminate, consequently it is impossible to judge of their validity except in terms of expediency.

The Adjustment Factor. In the search for tendencies toward consistent relations between scores, it was observed that the scores of many children when plotted fall on a straight line. (Table XXI, Fig. 5.)

TABLE XXI
Straight Line Adjustment Curves

Handwriting Scores of Children, Selected from Grades 3-7, for Whom the Scores in Fast, Normal, and Slow Handwriting Tests Fall in Approximately a Straight Line

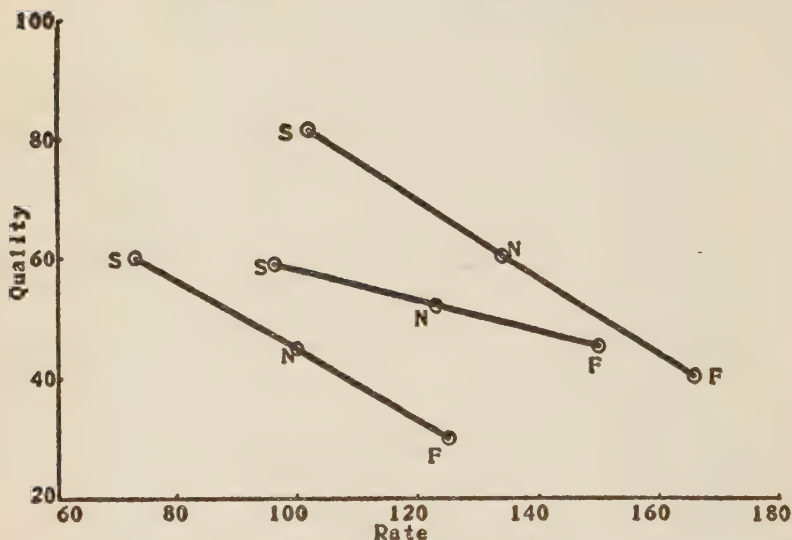
Individual	Age	Grade	Fast		Normal		Slow		Mid-Point		Angle of Slant	Amount of Adjustment
			R.	Q.	R.	Q.	R.	Q.	Fast	Slow		
1	18	19	57	33	50	56.5	50	57	54	45	16°	25
2	21	22	47	36.5	44	45	34	66	40	51	24°	32
3	18	17	31	24	20	46.5	19	49	25	36	26°	28
4	18	18	46	38	42	44.5	39	49	42	44	32°	13
5	22	22	78	29	54	61.5	45	72.5	62	51	37°	55
6	21	23	78	37.5	54	71.5	48	77.5	63	58	37°	50
7	24	22	74	35	69	41	54	59	64	47	40°	31
8	19	20	60	53.5	45	65.5	42	69	51	61	49°	23
9	21	21	74	43.5	57	56.5	45	65.5	60	54	53°	36
10	21	22	69	64	51	77	50	77.5	60	71	55°	23
11	19	21	68	67.5	54	77.5	45	83	56	75	56°	28
12	25	27	78	72	64	78	52	84.5	65	78	64°	29
13	23	22	76	82	60	89	53	93	64	88	64°	25
14	22	22	88	31.5	52	50	38	56	63	44	64°	56
15	23	23	84	47	54	57.5	40	65	62	56	68°	48
16	21	19	47	28	28	35	30	34.5	38	31	69°	18
17	20	21	74	58	60	63.5	55	64.5	64	61	71°	20
18	21	21	59	61	48	65	32	69	46	65	74°	28
19	23	23	66	92	36	94	48	93.5	57	93	83°	18
20	18	18	38	66	26	59	31	61.5	34	64	123°	8

R = rate

Q = quality

FIGURE 5

Straight Line (Pure?) Adjustment Curves



Graph of scores of three adults in three handwriting tests. N means under normal or standard conditions, F means fast, S means slow. These records are similar to those of the children in Table XXI.

That is, for these children, a definite relation seems to exist between rate and quality. As one increases, the other decreases, and vice versa. Variations in score thus represent adjustments the children make to the changing conditions. Accordingly such lines will be called "adjustment" lines. Just as the driver of an automobile adjusts the positions of the firing control and of the fuel control to the traffic situation, so apparently ability adjusts itself to the various test situations, expressing itself as high rate and low quality when the situation demands speed, and as low rate and high quality when the situation demands precision, or accuracy of movement.

Although the selected lines in the figure make various angles with the axes for rate and quality, they suggest that there may be a general tendency toward a definite angle of slant. Accordingly the angle of slant for each adjustment line (the line joining

the rate-quality position in the Fast test to the corresponding position for the Slow test) was computed; also the length of the line between these two positions was determined.⁴⁴

TABLE XXII

Relation Between Quality of Handwriting and (1) the Angle Between the Adjustment Line and the Quality Axis, and (2) the Amount of Adjustment Made from Fast to Slow Writing—Grades 3-8

Quality of Writing Ayres	Number of Cases		Median Angle of Adjustment Degrees		Median Amount of Adjustment Units	
	1921	1922				
95	1	1	80	88	45	24
90	6	4	67	80	18	17
85	5	16	62	65	31	22
80	18	29	57	68	25	26
75	34	53	52	58	40	23
70	59	79	52	70	38	27
65	80	107	50	63	40	27
60	112	105	49	62	37	27
55	95	139	44	53	42	29
50	109	146	49	47	40	30
45	112	131	50	43	32	28
40	116	101	53	51	34	22
35	104	102	56	53	26	27
30	81	67	52	69	19	17
25	51	46	61	68	17	12
20	18	23	57	93	19	9
15	10	1	67	73	15	10
10	9	—	82	—	22	—
Total	1020	1150	Low Point		High Point	
	Median Quality		44		42	
	50.4	53.6	Degrees		Units	

Probable true mid value for quality allowing for larger number of children in lower grades in 1921, 55 Ayres.

Probable corresponding angle of slant, 45°.

Probable corresponding adjustment, 42, or a change of 30 letters per minute in rate and a change of 30 points in quality.

⁴⁴Formula:

$$\text{Tan} = \frac{R_F - R_S}{Q_S - Q_F}$$

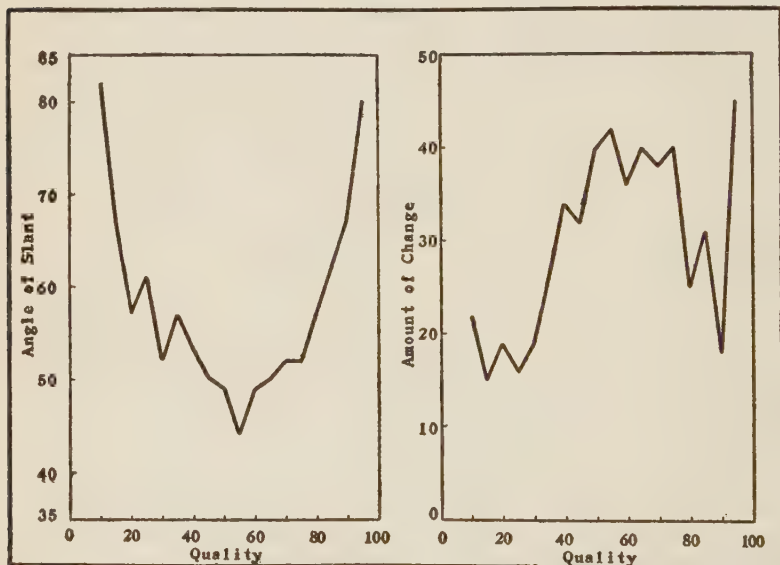
$$L = \sqrt{(R_F - R_S)^2 + (Q_S - Q_F)^2}$$

where Tan = the tangent of the angle of slant made by the adjustment line with the quality axis, R = rate, Q = quality, F = fast, S = slow, L = length of line of slant.

From the tabulation of these data, certain tendencies begin to be evident (Table XXII, Fig. 6). For writers of average quality,

FIGURE 6

Relation of Angle of Slant and Amount of Adjustment to Quality of Writing



The figures show that both the poor and the able writers make adjustments in rate only and that the amount of adjustment is small. The angle of slant changes toward 45° and back again as quality changes from low to average (55 Ayres) to high. Amount of adjustment increases and decreases correspondingly.

the average angle of slant is approximately 45°, although poor writers and good writers not only vary from this angle but are able to make but small adjustments. Repetition of the tabulation for the tests given a year later than the first gave the same general relationships, but a slightly narrower range of variation.

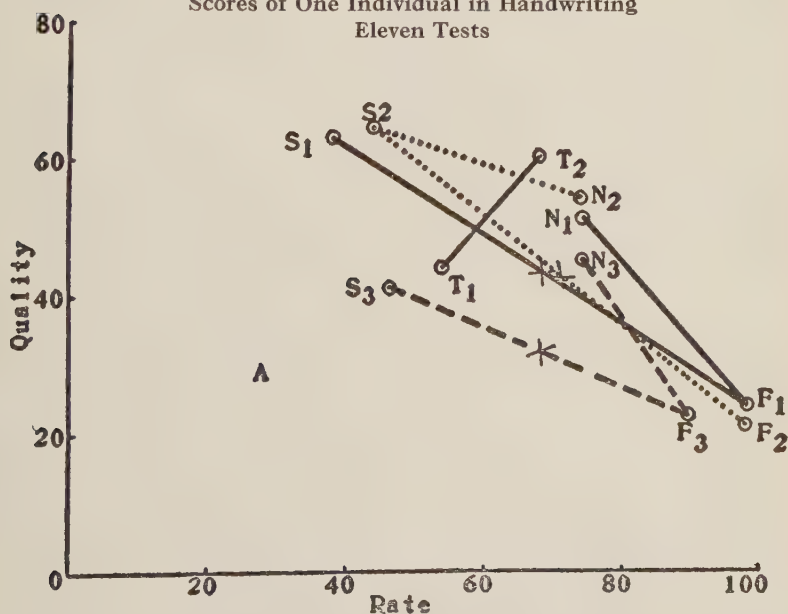
However, the outstanding characteristic of these results is their variability. However truly the adjustment theory may account for variations in rate and quality along a 45° line, most of the variation was not of this character. (Table XXIII,

TABLE XXIII
Scores of a Typical Individual in a Series of Eleven Tests in
Handwriting Within Ten Days

Test No.	Examiner	Conditions	Scores		Sum of Rate and Quality	Midpoint Slow—Fast	
			Rate	Quality		Rate	Quality
1	Teacher	Standard	54	44	99		
2	Student A	Standard	74	51	125		
3	Student A	Fast	98	24	122	68	44
4	Student A	Slow	38	63	101		
5	Student B	Standard	74	54	128		
6	Student B	Slow	44	64	108	71	43
7	Student B	Fast	98	21	119		
8	Student C	Standard	74	45	119		
9	Student C	Fast	90	22	112	68	32
10	Student C	Slow	46	41	87		
11	Teacher	Standard	68	60	128		

FIGURE 7

Scores of One Individual in Handwriting
Eleven Tests



NOTE: N means test given under normal or standard conditions; F, test given emphasizing speed of writing at expense of quality; and S, the test given emphasizing quality at expense of speed. 1, 2, 3 mean first, second, and third examiners. T₁ and T₂ means teacher's own test before and after the visit of the three special examiners. The tests used were identical throughout the ten-day period.

Fig. 7.) For instance, the scores of one individual in eleven tests within ten days range from 38 to 98 in rate and from 64 to 21 in quality. Even the sums of the various rates and qualities, which would be constant if the adjustment line were a 45° line, vary from 87 to 128. Nor is there any consistent pattern in graphs of individual scores, as the curves in Table XXIV, Figure 8, make evident. Therefore, although adjust-

TABLE XXIV

Scores of Four Individuals in a Series of Eleven Tests in Handwriting Taken at the Same Time and Under the Same Conditions as Those Given in Table III

Test	Conditions	Individuals							
		B		C		D		E	
		Rate	Quality	Rate	Quality	Rate	Quality	Rate	Quality
1	T	48	38	54	69	62	73	54	47
2	N	48	42	58	70	64	74	58	32
3	F	54	28	54	64	76	57	58	30
4	S	44	37	52	80	56	74	50	48
5	N	54	37	56	67	62	65	68	48
6	S	58	41	58	80	54	79	56	46
7	F	54	28	68	61	90	56	72	40
8	N	46	39	56	59	68	61	68	28
9	F	44	24	68	60	76	48	62	22
10	S	42	32	42	62	58	73	42	29
11	T	62	43	64	68	68	73	74	29

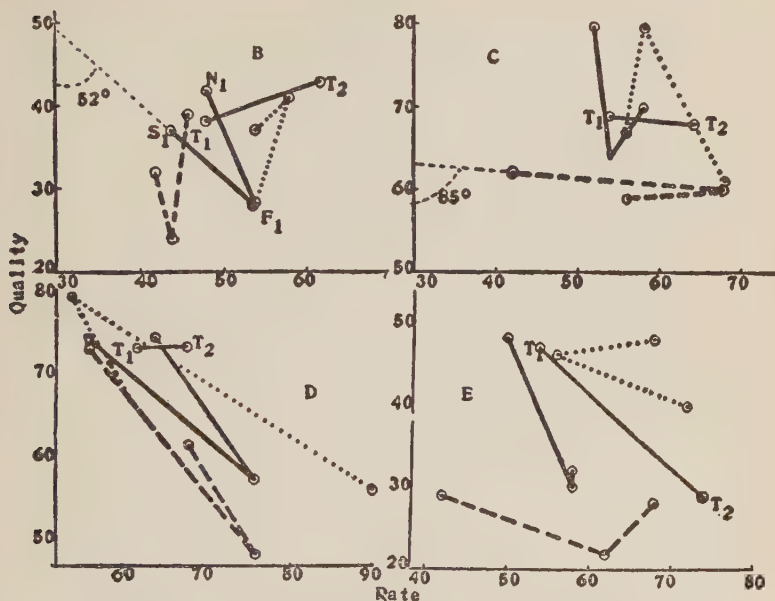
ment may account for part of the variation, it is not adequate to account for it all.

The Effort Factor. A factor which evidently influences response greatly is effort.⁴⁵ A subject may refuse to write anything at all, or, at the other extreme, he may put forth the maximum effort of which he is capable. In the first case, his score would be zero; in the second, it would be greater than the score which represents ordinary or average effort.

Accordingly, two experiments were devised to test the effect of effort; (1) a special appeal from the teacher, and (2) a money

⁴⁵Thorndike, E. L. "The Measurement of Intelligence," *Psychological Review*, 31:228, May (1924).
 Thorndike E. L. and Curtis, S. A. "Correction Formulae for Addition Tests." *Teachers College Record*, XXI; 9, January (1920).

FIGURE 8



Individual Variation in Response to Changing Conditions

NOTE: The curves of four other individuals tested at the same time and under the same conditions as individual A. For B and C the angle of slant of the adjustment line is shown. The lengths of these adjustment lines are B, 13; C, 26. N, F, and S have been omitted except in the case of the first series of tests of B. Note that the order in the second trial (dotted line) was normal, slow, fast.

reward. The teacher gave three tests, normal, fast, and slow, as nearly as possible under ordinary conditions. Then she told the children that she had heard that a strange examiner could come in and get better results than the regular teacher, but did not believe it was true; that she would be "dreadfully mortified" if anything like that happened to her; that she was going to repeat the three tests so that each child might have an opportunity to show how much he cared for her by doing the utmost of which he was capable. She then repeated the three tests. In the afternoon of the same day, she repeated the three tests again, this time under standard conditions, to see whether or not the effect of the special effort held over.

The next day, a special examiner came to the room and gave the three tests, as the teacher had done. Then she placed a bowl of dimes on the teacher's desk and promised a dime to each child that would beat his former record in **both rate and quality**. The children were much intrigued, and after asking numerous questions as to whether or not they could keep the dimes to spend as they saw fit, and what the conditions were under which they would receive the dimes, they took the three tests again, very evidently putting forth much effort. After each test, the examiner went to each child, who on count, found he had written a larger number of letters than before and who judged his quality of writing to be better. If she concurred in his judgment, the dime was paid on the spot. Cases of doubt were referred to the teacher and to classmates for decision. In the afternoon, the examiner returned and repeated the three tests under regular conditions without money reward, asking the children to write as they usually did. Two classes were tested in this way, and complete records secured from 73 children, 34 girls and 39 boys.

In one class of 34 children, eight were successful in all three trials, earning 30 cents apiece, seventeen children earned 20 cents, and there was no child who did not succeed in earning at least one dime. The paying out of these 67 dimes as the tests proceeded served as a continuous stimulus to effort.

Judgments made in class at the time were necessarily imperfect judgments. But a study of the records after the papers had been carefully scored, shows that both the teacher's appeal and the money payment operated to raise the performance of three-fourths of the children, and that the money appeal on the whole called out the greater effort. (Table XXV.) The magnitude of the response is of some interest. (Table XXVI.)

The teacher's appeal produced a slightly greater absolute change in performance than the money reward in every case except the last test. This probably means that the effect of the teacher's appeal wore off as the tests continued, while the repeated payment of money acted to renew the stimulus. Also, it must be remembered that a slightly larger per cent of the

TABLE XXV

Comparison of the Effects of Two Stimuli on Effort

Number and Per Cent of Children Making Higher Scores under the Stimulus (1) of Teacher's Personal Appeal and (2) of a Money Reward for Success. Total cases 73.

Stimulus	Number of Children																			
	Increasing in Both Rate and Quality				Increasing in ² Either Rate or Quality				Increasing in ³ Combined Score				Total No. Achieving Increase				Per Cent Achieving Increase			
	N ¹	F ¹	S ¹		N	F	S		N	F	S		N	F	S	N	F	S		
Money Reward	24	37	22		7	2	3		23	18	30		54	57	55	74	78	75		
	26	23	14		7	3	7		19	26	19		52	52	40	71	71	55		
Difference % of Difference ⁴	-2	14	8		0	-1	-4		4	-8	11		2	5	15	3	7	20		
	-3	19	11		0	-1	-5		5	-11	15		3	7	21	4	10	27		

¹N = normal, F = fast, S = slow.

²The other characteristic remaining constant.

³The gain in one overbalancing the loss in the other characteristic.

⁴Base 73.

TABLE XXVI

Change Measured from the Teacher's Normal Tests to Teacher's Appeal Tests and from Examiner's Normal Tests to Examiner's Payment Tests. Cases = 73

Magnitude of Changes Computed by the Formula.

$$\text{Change} = \sqrt{(\text{Difference in rates})^2 + (\text{Difference in qualities})^2}$$

Comparison of magnitude of effects of Teacher's Appeal and Examiner's money payment.

Magnitude of Change	Teacher's Tests			Examiner's Tests		
	Normal	Fast	Slow	Normal	Fast	Slow
Over 50	1	1			1	1
45	1	1			1	0
40	0	1			1	1
35	1	2		0	1	2
30	1	6	1	2	2	2
25	6	10	3	0	4	6
20	7	9	7	3	6	15
15	6	15	17	10	21	10
10	24	11	14	27	17	17
5	16	13	21	18	15	14
0	10	4	10	12	4	5
Median	12.2	17.8	12.0	11.2	16.2	15.2
Equivalent ¹ Change in Ability	8.6	12.6	8.5	7.9	11.5	10.7

¹Equivalent Change in Ability = median change $\div \sqrt{2}$. That is, a change of 12.2 is equivalent to a simultaneous change of 8.6 letters in rate and 8.6 points in quality.

children responded to the money payment. In both cases, however, fatigue, nervousness, or other factors, operated to make certain children respond with largely decreased scores, so that the results as a whole, while proving that incentives which stimulate effort modify performances, show also that the effect of the stimulus cannot be foretold without a knowledge of the other factors operating. A child who is already operating with the optimum effort for maximum efficiency, may seriously interfere with the perfect functioning of his mechanism by putting forth more effort.⁴⁶ However, I regard the evidence

⁴⁶Illustrations of fluctuations in efficiency caused by the variations in the degree of successful coordination of the factors involved are given in

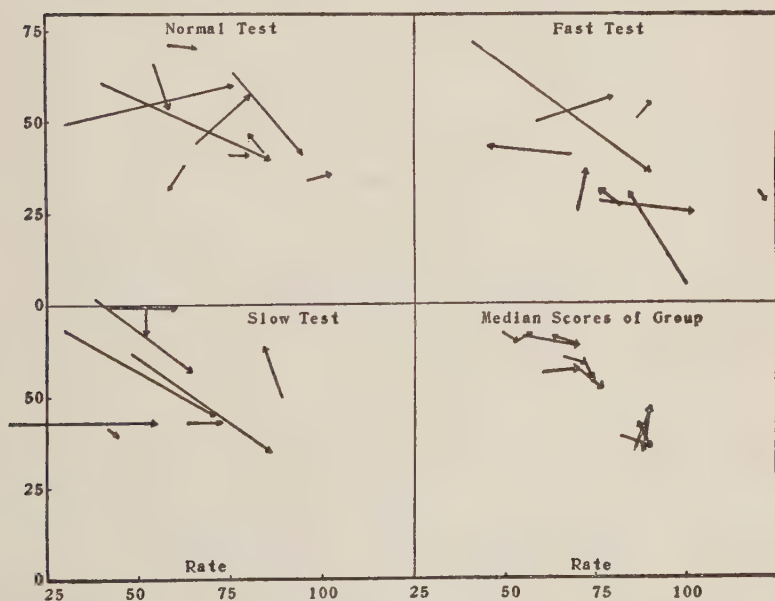
Thorndike, E. L. and Curtis, S. A. "Correction Formulae for Addition Tests," *Teachers College Record* XXI:10. January (1920).

presented as proof that effort is a factor affecting performance in handwriting. The next step will be to determine the nature of the effect of effort.

Effect of Effort. A study of the individual changes reveals an almost unbelievable variation. (Figure 9.) Ten cases have been chosen as illustrations from each test—all boys. The records for girls show similar characteristics. It is evident that the factor of individual difference plays too large a part to permit generalization from these changes.

FIGURE 9

Individual Variation in Response to a Money Stimulus



Normal, fast and slow refers to the tests made using a money reward as a stimulus to effort. Each arrow represents the change in score of an individual child. The end of the arrow shows the rate and quality in the test without the stimulus; the point of the arrow indicates the score made in response to the stimulus. The length of the arrow represents the amount of change and the slant and direction of the line, the quality of the change. The arrows in the lower right-hand rectangle indicate in similar fashion the median changes in the scores of the entire group.

Individual variation may be removed by averaging (Table XXVII, Fig. 10). The adjustment lines joining the average

TABLE XXVII

Effect of Effort Upon Class Adjustment Lines

Comparison of changes in median scores in handwriting of a group of 73 sixth grade boys and girls under the stimulus of special incentives: (1) the personal appeal of the teacher and (2) money payment for success.

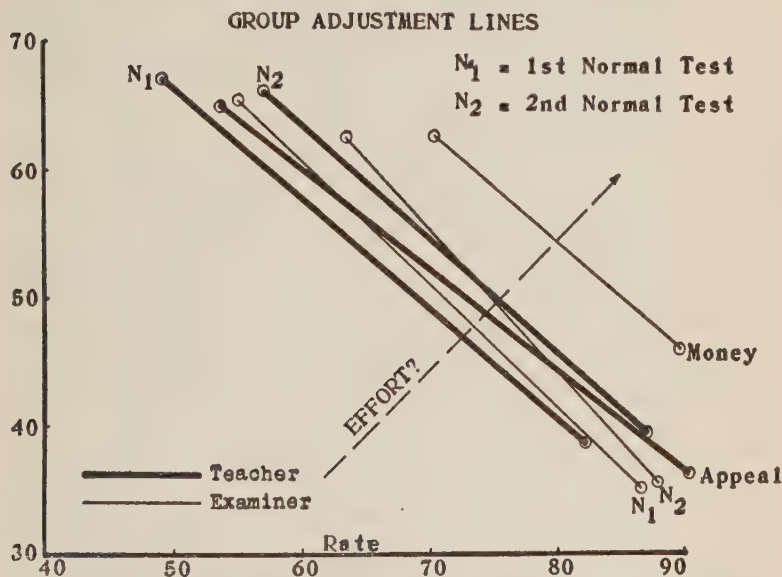
Tests	Tests						Scores		
	Normal		Fast		Slow		Sum of Rate and Quality		
	Rate	Qual- ity	Rate	Qual- ity	Rate	Qual- ity	Normal	Fast	Slow
Teacher's									
Normal.....	59.8	55.9	82.2	38.7	49.0	67.2	115.7	120.9	116.2
Appeal.....	70.2	57.3	90.3	36.2	53.5	65.2	127.5	126.5	118.7
Normal.....	77.1	52.5	87.0	39.5	56.9	66.2	129.6	126.5	123.1
Examiner's									
Normal.....	66.5	59.7	86.4	35.2	55.0	65.6	126.2	121.6	120.6
Money.....	71.9	59.0	89.7	46.0	70.2	62.7	130.9	135.7	132.9
Normal.....	73.6	54.2	87.9	35.4	63.3	62.7	127.8	123.3	126.0

group scores in the fast and slow tests for each of the six trials, are approximately 45° lines, but in relation to each other are displaced along a line at right angles to the adjustment lines, as indicated by the dotted line in the figure. Accordingly, the scores of each of the 73 individuals were graphed and, without exception, evidence of similar displacements was obtained. Therefore, the hypothesis is advanced that variations in performance are caused in the main by two factors, adjustment and effort. According to this hypothesis, each change in performance from one test to another, where no change in ability has taken place, is primarily the resultant of these two factors.

For instance, if we graph all the available scores for an individual who was present for the entire series of tests that were given, the points representing the scores outline vaguely a surface that appears to be an ellipse (Figure 11). If the number of tests had been three hundred instead of thirty and the conditions and incentives had been more varied, the figure might

FIGURE 10

Displacement of Group Adjustment Lines Under Effort



The adjustment lines join the median group scores in the fast to the median group scores in the slow test in each case. N = Fast-Slow tests under normal conditions: that is, without special stimulus. Appeal = Fast-Slow tests under stimulus of teacher's appeal; money = Fast-Slow tests under stimulus of money reward. The direction of displacement under effort is indicated by the arrow.

have been more clearly defined. For the present, however, the field of variation will be assumed to be an ellipse. The adjustment line forms the major axis, and the effort line the minor axis. These are at right angles to each other, and both form angles of 45° with the axis. Any performance which varies from another can be immediately resolved into the two factors, effort and adjustment, by the formula⁴⁷

$$\text{Adjustment} = \sqrt{.5} [(X_2 - X_1) - (Y_2 - Y_1)]$$

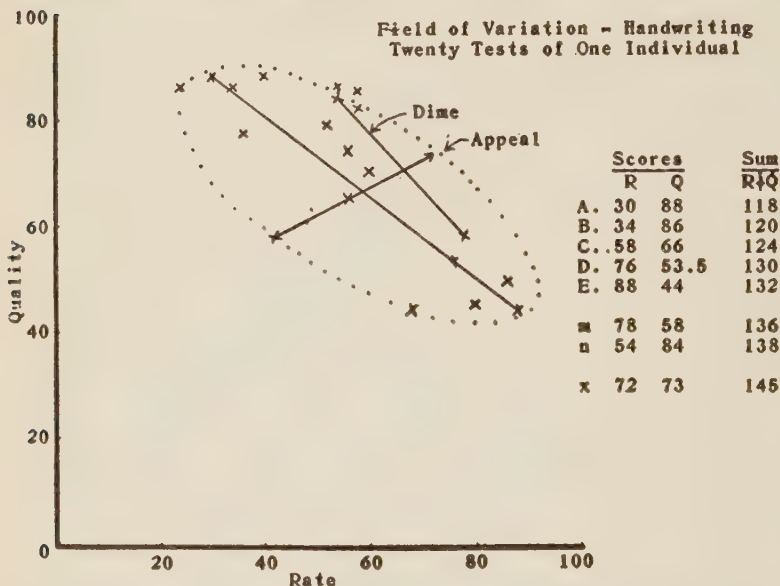
$$\text{Effort} = \sqrt{.5} [(X_2 - X_1) + (Y_2 - Y_1)]$$

⁴⁷See Appendix for derivation of the formula.

FIGURE 11

Field of Variation for One Individual

VARIATION IN PERFORMANCE CAUSED BY CHANGE IN FACTORS AFFECTING
FUNCTIONING OF ABILITY



Each cross represents one test. The dotted ellipse outlines the (theoretical) field of variation. The longest line is the adjustment line; the line across its center, the effort line. The lines of adjustment under a money reward and the teacher's appeal are indicated.

where X_1 Y_1 are the rate and quality of the position to which the second performance is referred, and X_2 Y_2 the rate and quality of the second performance.

If now it should become possible to measure effort objectively and directly, the hypothesis could be proved or disproved. Therefore, a vital problem that must be considered is

Problem 6. Direct Measurement of Effort

Unfortunately, no direct method of measuring effort is known at the present time. Therefore, the sixth problem that must be solved before the reasons why children succeed in school can be

more completely determined is the problem of finding a direct measure of effort.

Laws of Variation. In the meantime, scores in the tests may be handled "as if" the hypothesis were true. On this basis, it is possible to state two laws of variation in performance, (1) the law of adjustment, and (2) the law of effort. Also, it is possible to obtain a hypothetical measure of effort from performances which vary from each other in some other direction than along the adjustment line.

The law of adjustment is that rate and quality, in a rate test where equivalence of units between rate and quality has been established, vary inversely, their sum remaining constant. This constant sum is, therefore, the single measure of ability sought. It is neither rate nor quality, but ability. Its development and significance will be discussed in later chapters. It is enough to say here that to be comparable, all rate and quality scores in rate tests should first be reduced to equivalence, probably in terms of relative changes in rate and quality when adjustment takes place in response to a change in stimulus. The shape of the field of variation will be an ellipse and the adjustment line a 45° line, only when the units have been equated. Special laws might be developed for each type of relationship between the rate and quality units. Whatever the form of the field of variation under the adjustment-effort hypothesis, ability will be represented by the formula

$$A = C_1R + Q$$

or

$$R + C_1Q$$

where A = ability, C_1 = a constant, R = rate score, and Q = quality score, and not by a formula expressing a multiplicative relationship as long as the relationship is rectilinear.⁴⁸

The law of effort is that both rate and quality vary directly with effort and fluctuations in the sum of the equated rate and quality scores are measures of differences in effort. That

⁴⁸For instance, the formula for Gates' results (page 116) is $A = .065R + Q$. By this formula the ability scores for the Scarborough averages become 13.5, 13.6, 13.3 respectively, scores which are just as constant as those given by Gates' formula and which have the added advantage of causal explanation. The reason for the large reduction in the weight given rate scores is not the unimportance of speed as Gates implies, but the relative size of the units used to measure rate and quality. Roughly, one Thorndike unit of quality is equal to 16 units of rate.

is,

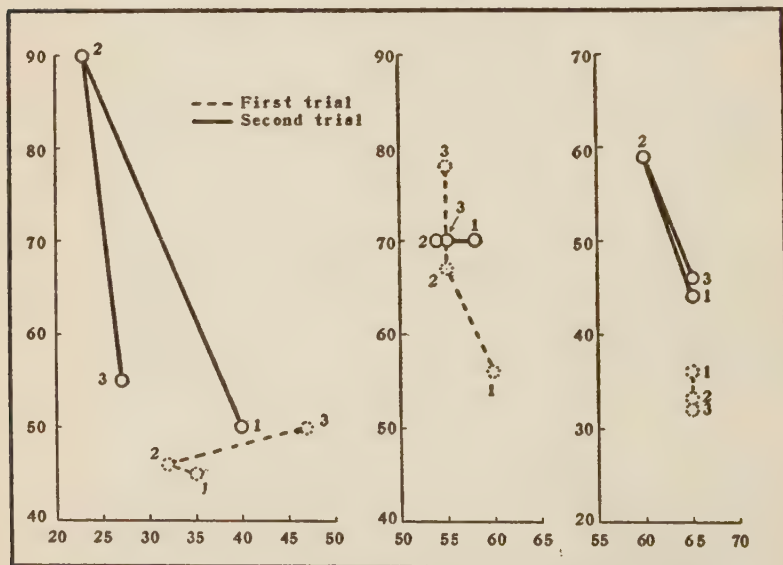
$$E_1 - E_2 = [(R_1 + Q_1) - (R_2 + Q_2)]$$

where E = effort, R = rate score, and Q = quality score.

It might be thought that children would show consistent patterns of variation from day to day, but such is not the case. (Figure 12.) While these are selected cases, chosen to represent extreme fluctuation, many such variations occur. They are but graphic representations of the changes taking place within the individual, and are of the utmost significance both for testing and for teaching. For very evidently the re-

FIGURE 12

Individual Variation in Handwriting Tests on Different Days



Tests 1, 2, 3 were given under "normal," "slow," and "fast" conditions. The solid line represents test given one day; the dotted lines, tests given one or two days later. The diagrams are selected to show extreme variations in behavior.

sponse a child makes to an external situation is determined both by the situation and by his internal state. Recognition of this fact is leading to the abandonment of the old uniform mass instruction in favor of individualized methods of teach-

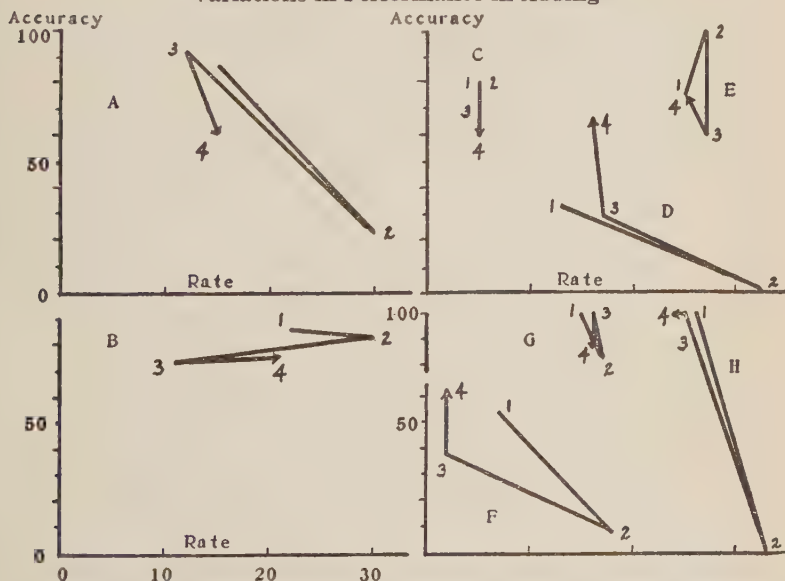
ing.⁴⁹ The need of the hour in the measurement field is for measuring instruments that will reveal inner conditions and enable adjustments of the external situation to be made accordingly.

Similar Variation in Other Rate Tests

The question naturally arises, "How far do the conditions revealed for handwriting hold for rate tests in other subjects?" The answer is that while it has proved more difficult to secure comparable records of variation, evidence has accumulated that

FIGURE 13

Variations in Performance in Adding



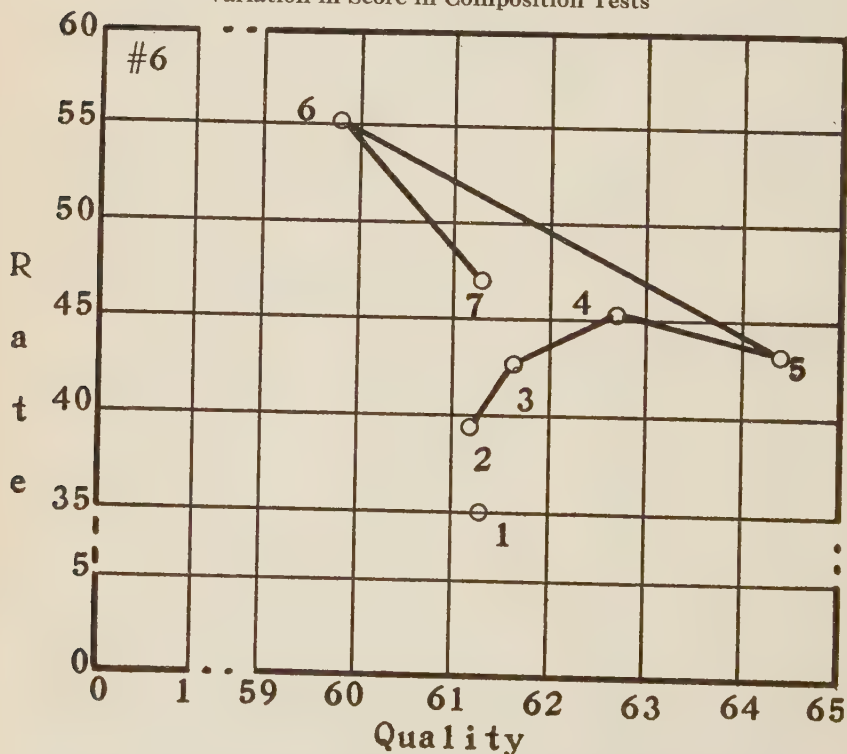
Records of eight individuals selected to represent the various types of changes in individual performance which take place when the conditions of testing are varied. Tests 1, 2, 3, 4 were given in order. The first was under normal conditions, the second, a "fast" test, the third, a "slow" test, the fourth under the same conditions as the first. All graphs are drawn to scale, but the rate axes for Individuals C, D, E, F, G, H have been shifted as necessary to bring the curves closely together. Consequently no rate scale is shown. Individual A is a typical individual in the sense that the 2-3 line is in conformity with theory. Individual B shows a response of an opposite character. D, F, H, exhibit adjustment lines in conformity with theory but on different accuracy levels. Individuals C, E, G, represent special forms. Note the individual E ended where he began.

⁴⁹See "Adapting the Schools to Individual Differences." *Twenty-fourth Yearbook of the National Society for the Study of Education*. (1925).

variation exists, and, in the main, exhibits the same general tendencies.

Addition. For instance, a series of rate tests was given in addition.⁵⁰ Some of the results show precisely the same type of adjustment curves, while others show the type of erratic variation that would be expected from extreme changes in

FIGURE 14
Variation in Score in Composition Tests



Rate and quality scores in composition tests under various conditions. Test 1 and Test 2 were compositions on assignments rated by Hudelson⁵¹ as typical. Test 3 was Hudelson's maximal assignment. Test 5 was a test given under conditions considered by the writer to emphasize quality (slow) while Test 6 emphasized speed (fast). Test 7 was a typical assignment given as a check test. The line 5-6 is an adjustment line corresponding to similar lines in handwriting.

⁵⁰See Appendix

⁵¹Hudelson, E. "English Composition," *Twenty Second Yearbook, Part I*, National Society for the Study of Education (1923) pages 77-99.

effort. (Figure 13.) It will be remembered that the adjustment-effort hypothesis was first developed in this field.

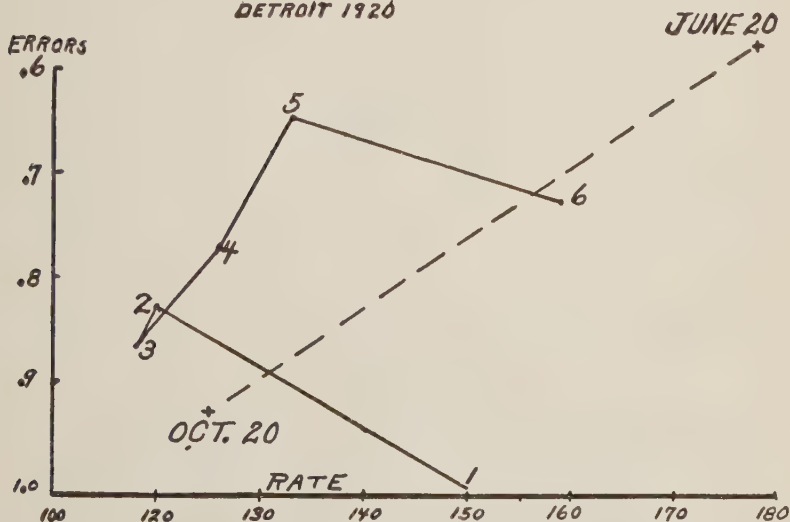
Composition. A series of tests in English composition given with various emphases on rate and quality yielded precisely similar results. (Figure 14.)⁶²

Typewriting. Attention will be called a little later to similar results in reading,⁶³ while records from a class in typewriting⁶⁴ prove that the same general situation exists in that field. When all the evidence is reviewed, and no conflicting data found, conviction grows that whatever the final form the theory may take, the types of factors here designated—adjustment and

FIGURE 15

Variation in Rate and Quality of Typewriting

**PROGRESS IN TYPEWRITING
BY WEEKS--COURSE 2
DETROIT 1920**



Rate = number of strokes per minute, quality, the number of errors per 100 strokes. The crosses indicate class scores at dates shown. The solid line represents the record of one individual in various tests between the two dates.

⁶²From an unpublished study by Rankin, P. T. and Courtis, S. A., reported before the National Society for the Study of Education at Cleveland, Ohio, meeting. (1923).

⁶³See pages 143-147.

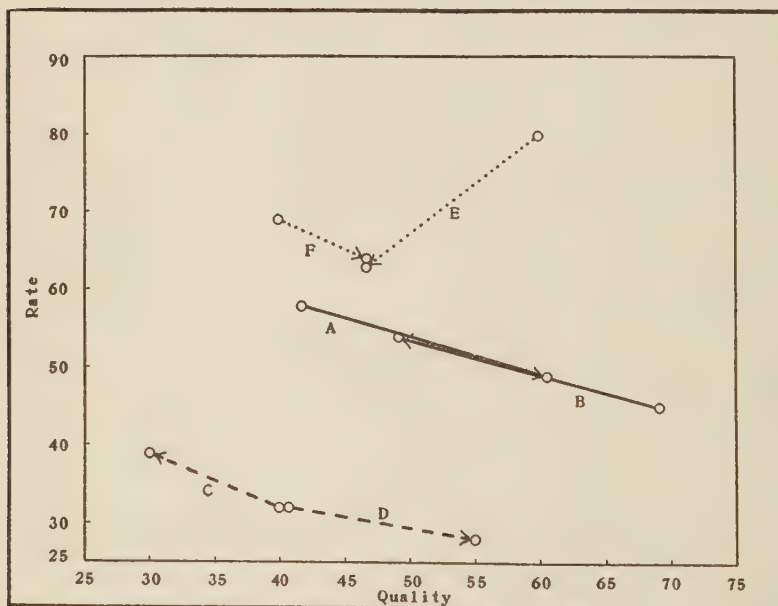
⁶⁴From an unpublished study by E. G. Blackstone, now of the University of Iowa, then of the Detroit Public Schools. (Figure 15.)

effort—must be brought under control before scores in performance tests can be correctly interpreted.

Dangers in Conventional Interpretation. The full significance of the play of the adjustment and effort factors upon scores is best shown graphically (Figure 16). When children are given a test under standard conditions, the response of each child, as has been shown, is the resultant of three major variables,

FIGURE 16

Illustrations of Lack of Comparability in Handwriting Scores



A-B, C-D, and E-F are three pairs of individuals selected to represent typical cases. Their scores (not comparable) in a standard test in handwriting are represented by the tails of the arrows; the centers of their fields of variation (comparable) are represented by the points of the arrows. A, who on a single test would be judged less able than B, is really more able. C, who on a single test would be judged equal in ability to D, is really less able. E who on a single test appears much more able than F, is really the equal of F in ability.

ability, adjustment, and effort, the relative amounts of which differ from child to child. Consequently, one child's response represents the center of his field of variation, another child puts forth nearly his maximum effort, a third child makes an adjust-

ment in terms of speed. The resulting scores, to the degree that they are not comparable, misrepresent the relative abilities of the children.

For instance, children A and B in the figure seem to differ widely in ability, and B would ordinarily be rated much better than A. He appears to excel by ten points in rate and thirty points in quality. But B has written his best and A his fastest. If their entire fields of variation are explored and their relative abilities judged from the position of the centers of their fields, that is, by measures of ability which represent the same conditions in the behavior of the two children, A is more able than B. C and D represent the opposite case. By a single test, they appear to be of equal ability; by comparable measures, D is much more able than C. F and E represent the case of two children who by a single test might be judged to be very different, but who by comparable measures are of equal ability.

These are, of course, selected cases, but the tables previously presented have made it plain that effects of this sort are operative in every score, and that the magnitudes of the effects are great enough to obscure seriously both correlations and relationships accordingly. Correction of raw scores to standard conditions seems destined to play as large a part in educational science as, by way of illustration, the correction of the volume of a gas to standard conditions plays in chemistry.

Performance Scales. In contrast with rate tests, performance scales have been developed to aid in determining how far the mental or educational development of an individual has progressed. They do not measure skill, but "power" or "level" or whatever characteristic of ability it is which enables the mature to achieve tasks which are impossible to the immature.

Since the goal of the use of a performance scale is differentiation on the basis of development, performance scales are either untimed or the time interval is long enough so that each individual has a chance to try each element, and to do it if he can. Theoretically, score is not based upon amount done, although often computed by counting methods, but indicates the degree of difficulty of the most difficult element the subject can handle

successfully. The fundamental problem in the construction of performance scales is thus the determination of the degree of difficulty of the elements.

Method of Construction. The method generally accepted is to scale the elements in variability units based on the percentage failing.⁵⁵ The result obtained depends upon the variability of the group, and recent efforts have been directed toward standardization of the group.⁵⁶ There are many problems here in need of solution, but so much has been written by the various authorities pointing out the difficulties and so little evidence is available, that the problems of determining difficulty and locating zero-points will be passed over. It should be noted in passing, that these conditions, while they do not absolutely prevent the determination of law, make it more difficult.

The point to which attention will be paid is a fundamental one from the point of view of analysis. In a performance scale, each element differs in complexity from all the others not in its own class, and is really a separate test. A performance scale is, therefore, merely a battery of tests, and by making the test long enough and providing enough time the extraneous elements are (theoretically) averaged out so that the final score represents pure development. The constancy of scores in performance scales is often contrasted with the variation in the scores from rate tests, to the latter's disparagement. I shall now present evidence to show that skill and all the factors of adjustment and effort operate in performance scales and raise the question as to whether under the circumstances scores in performance scales can be accepted at their face value.

Appraisal: Spelling. In a certain investigation in spelling,⁵⁷ a performance scale spelling test, composed of a pair of words for each column of the Ayres Spelling Scale (Figure 17), was

⁵⁵Monroe, W. S. *Introduction to the Theory of Educational Measurement*. (1923.)

⁵⁶McCall, W. A. "A Proposed Uniform Method of Scale Construction." *Teachers College Record* XXII:31, January (1921).

⁵⁷From an unpublished study by Marie I. Rasey, Head of Department of Educational Research, Detroit Teachers College.

FIGURE 17

Illustrative Selections from a Performance Scale in Spelling
Made by Choosing Two Words from Each Section of Ayres' Scale

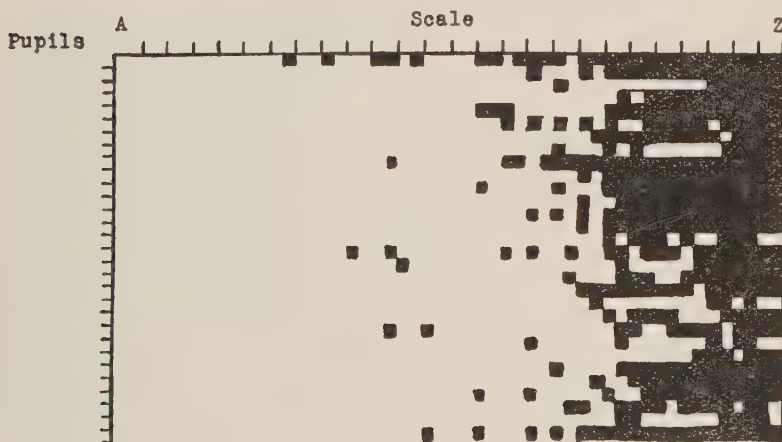
Section of Scale	Word Chosen	Section	Word	Section	Word
A	do	K	brother	W	practical
A	me	K	coming	W	appreciate
B	go	L	country	X	immediate
B	on	L	morning	X	committee
C	can	M	pretty	Y	decision
C	run	M	recover	Y	principle
D	man	N	capture	Z	judgment
D	top	N	teacher	Z	recommend

Note that the words steadily increase in difficulty from first to last.

FIGURE 18

Scores Made by Individual Pupils in Performance Test in Spelling

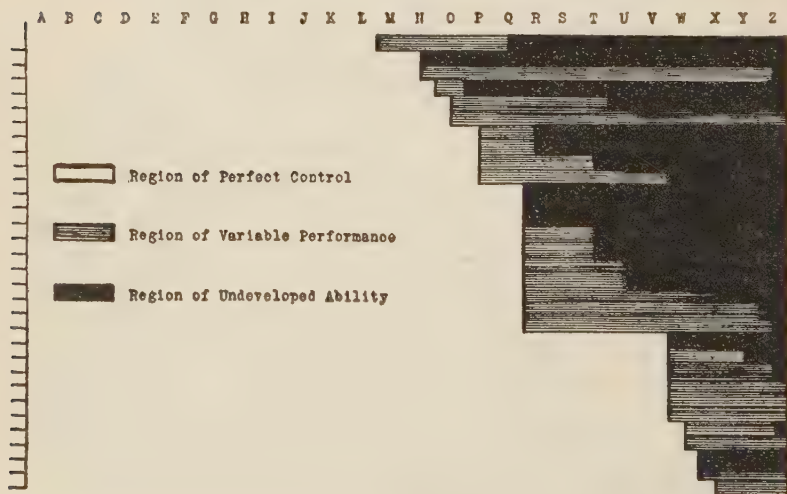
RESULTS FROM AYRES SPELLING SCALE



Scale across the top represents the sections of the Ayres Scale, each division representing two words. Scale at the left represents a class of 30 children, each division representing one pupil. That is, each small square represents one word spelled by one pupil. If the spelling is right, the square is white; if the spelling is wrong, the square is black. On the easy portion of the test, all squares are white, indicating a region of perfect control; then comes a region of variable control where black and white alternate; finally there is solid black indicating a region of no control.

given to a class of thirty children. The results are shown graphically in Figure 18. For most children, there were many words in the easier part of the scale that all spelled perfectly. Ultimately, however, words were reached of such difficulty that a given child would spell one of the pair for that level and miss the other. As progress is made through the list, the frequency of misspelling becomes greater, until finally none of the words are spelled correctly. That is, the three regions of perfect, variable, and no control are plainly marked. Moreover, the differences between the performance of different children suggests that the children are putting forth different degrees of effort and varying their performance in precisely the fashion in which performance has been shown to vary in handwriting. Some children spell perfectly every word they know. For them,

FIGURE 19
Variation in Spelling Control



The results of the previous figure rearranged to bring together in groups all those whose regions of variable control begin at the same point. The scale at the top represents sections of the Ayres Scale, the scale at the left represents pupils but in a different order from that of previous figure. The largest group all began to miss at words in Column R. The first three pupils in the group show no region of variable performance but pass at once from perfect to no control. The others in the group show regions of variable performance of different widths.

there is no region of variable performance. Other children are uncertain about the spelling of many words, missing and succeeding over a wide area. (Figure 19.)

Much evidence has accumulated which proves the existence of fluctuation in performance in spelling.⁵⁸ The conventional manner of dealing with the unreliability of spelling tests caused by this variation in performance is to increase the length of the test.⁵⁹ However, if variation in spelling performance is caused by fluctuations in effort and adjustment as in writing, then the same misinterpretation of scores which was there shown to be possible, would be possible in spelling. Performance scales serve a useful purpose in that they differentiate pupils of one type of performance from another, but they probably do not measure spelling ability in the sense in which the term measurement is ordinarily employed.

Reading. Thorndike has recognized the possibility of a single measure of ability which should be predictive of performance on whatever level of difficulty it operates. In taking his Reading Scale Alpha 2, the average child of the middle grade passes from easy paragraphs and questions about the paragraphs which he is able to answer perfectly, through more difficult paragraphs and questions, of which he is able to answer only a part correctly, to paragraphs and questions so difficult that he can answer none correctly. His ability is reckoned at that theoretical point on the scale at which, if there were a paragraph and questions about the paragraph of that degree of difficulty, he would have 80% of his answers correct. If a child has 90% of right answers on one paragraph, 70% on the next paragraph, and 60% on the following paragraph, it is possible to find from the table given by Thorndike, a single score which is taken as a measure of his ability.⁶⁰

At its face value, such a procedure represents ideal measurement. Actually, the resulting score may be a close approxima-

⁵⁸Cook, W. A., and O'Shea, M. V. *The Child and His Spelling*. (1914), pages 82-87.

Ashbaugh, E. J. "Variability of Children in Spelling." *School and Society*. (1919) IX:93-98. January 18.

⁵⁹Otis, A. S. "The Reliability of Spelling Scales Involving a Deviation Formula for Correlation." *School and Society*. (1910.) IX:716-22. November 11.

⁶⁰Thorndike, E. L. "An Improved Scale for Measuring Ability in Reading." *Teachers College Record*. (1915.) XVI:31.

tion to the truth. But as matters stand, there are three aspects of the situation which need investigation: (1) the validity of the scale, i. e., the relation of performance in taking the scale as a test to the functioning of the distinctive mechanism used in reading; (2) the relative difficulty of the scale units, or the validity of the statistical method by which difficulty was determined; and (3) the degree to which effort and adjustment operate to determine performance. My position is that the evidence submitted above makes it highly probable that a score in such a performance scale is only a partial measure of an intricate complex of abilities and factors, so that it is almost futile to hope to discover relationships between factors, even where decisive relationship exists, until methods have been devised by which the complex may be analyzed into its elements. This could be done by repeated testing of identical children with rate tests representing the various levels of difficulty, as soon as the parts played by effort and adjustment in rate tests are reliably determined. In the meantime, scores in performance scales are mere indices of differences in behavior, significant and useful, but probably not suitable for the exact determination of law.

On this point Thorndike says:

"The score obtained by using the instrument to measure an intellect is in present practice either a number representing a summation of credits and penalties, or, more rarely, a number representing the grade of difficulty of the tasks which the person can respond to with some assigned percentage of correct responses. . . . In neither case (even supposing the measurement to be a perfect representation of the person's abilities) can the numbers be taken at their face value. . . . The numbers 1, 2, 3, 4, etc., designating the scores made by individuals, do not represent a series of amounts of intellect progressing by equal steps. . . . We have then no right to add, subtract, multiply, or divide, with these scores . . . in the way that we do with . . . heights or weights. . . . The numbers designating the scores made by individuals are usually not even approximately related to any true zero point."⁶¹

Stanford Achievement Tests. An analysis of the responses made by children tested in this investigation to the various

⁶¹Thorndike, E. L. "The Measurement of Intelligence." *Psychological Review*. (1924) XXXI:221—2, May.

TABLE XXVIII

Variation of Performance in Performance Scales

Based on the responses of identical sixth grade children analyzed question by question.

PART I. Scores of 41 children tested with Stanford Reading Test 2, Form B, in 1923, and again in 1924. Only wrong (w) and omitted (o) answers shown. All others are right. For individual A, star indicates variation in performance.

Question Number	Individuals						Question Number	Individuals					
	A		B		C			A		B		C	
	1923	1924	1923	1924	1923	1924		1923	1924	1923	1924	1923	1924
1							41						
2							42	w	*				
3							43	*	o			w	
4				w			44	w	o			w	
5							45	w	o	w	w		
6							46	*	w				w
7				w			47						
8							48	*	o				
9							49	w	o	w	w	w	
10							50	w	*		w	w	
11	*	w	w	w			51	w	*	w			w
12							52						
13				w			53	w	*	o		w	w
14							54	w	w	w	o	o	w
15							55			o		o	
16							56	w	o	o		o	
17							57			o		o	
18							58	w	w	o		o	
19	w	*					59	w	w	o	w	o	w
20							60	w	w	o	w	o	w
21	*	w		w			61			o		o	
22	w	*			w		62	w	*	o		o	w
23							63	*	w	o		o	w
24							64			o		o	o
25							65	w	*	o		o	o
26							66			o		o	o
27	w	*		w			67	*	o	o		o	o
28				w		w	68	w	o	o	w	o	o
29			w		w		69	*	o	o	w	o	o
30					w		70	w	w	o	w	o	o
31							71	*	o	o		o	o
32							72	*	o	o		o	o
33			w				73	*	w	o		o	o
34							74	w	*	o	w	o	o
35							75	*	o	o	w	o	o
36					w		76	w	*	o		o	o
37			w				77	*	o	o	w	o	o
38							78	w	*	o		o	o
39						w	79	w	*	o		o	o
40	*	w	w				80	*	w	o	w	o	o
							Right	57	55	44	61	44	53
							Wrong	23	12	9	19	9	10
							Omitted	0	13	27	0	27	17
							Score 1923	34		35		35	
							Score 1924		43		44		44

TABLE XXVIII (Part) II)

Score of 47 Sixth Grade Children Tested with Standard Reading Test 2, Form B, Followed by Form A Two Days Later. Only wrong answers (w) and omitted answers (o) are shown. All others are right.

Question Number	Individuals						Question Number	Individuals					
	A		B		C			A		B		C	
	1923	1924	1923	1924	1923	1924		1923	1924	1923	1924	1923	1924
1							41	w					
2							42			w		w	
3							43		w		o	w	
4							44	w		w	o		
5							45			w		w	
6							46	w		w			w
7							47						
8							48	w					o
9							49	w		w		w	w
10							50	w		w		w	w
11							51		w				
12					w		52		w				
13							53				o		
14							54						w
15	w						55		w			w	
16							56				o		
17							57	w				w	
18				w		w	58		w		w		
19							59			w	w		
20							60			w	w		
21							61	w			o		
22							62	w		w			w
23							63		w	o		w	
24				w			64		o				
25							65		w	w			w
26							66			w	w		
27	w	w					67				o		o
28							68						o
29							69	w		o			o
30							70	w	w	o		w	o
31							71		w	o	o		o
32							72			o		w	o
33				w		w	73	w		o	w	w	o
34		w					74	w	w	o	w	w	o
35							75			o	w		o
36							76		w	o	w	w	o
37							77		w	o	w	w	o
38				o			78	w		o			o
39				o			79	w	w	o			o
40							80	w		o		w	o
							Right	62	64	57	60	63	57
							Wrong	18	15	11	11	17	8
							Omitted	0	1	12	9	0	15
							Score 1923	44		46		46	
							Score 1924		49		49		49
Right both tests.....									42		37		40
Right one test, omitted or wrong the other									28		31		23
Wrong or omitted both tests.....									10		12		17

performance scales which compose the Stanford Achievement Tests, makes clear both the extent to which the three regions of perfect, variable, and no control exist, and the degree to which the resulting score is an index of behavior rather than a measure of ability. (Table XXVIII.)

TABLE XXVIII (Part III)

Comparison of Extent of Fields of Control

Number and per cent of answers in different categories from test of 41 children a year apart, Form B, Stanford Reading Test 2, and of 47 children two days apart, Form B followed by Form A.

Categories	Number of Answers		Per Cent	
	Year	Two Days	Year	Two Days
Right in both tests (perfect control)	1780	2124	54	56
Right in first and wrong in second	232	370	7	10
Right in second and wrong in first	383	371	12	10
Tried in first, omitted in second	370	379	11	10
Tried in second, omitted in first	195	218	6	6
Wrong in both tests	148	169	5	4
Omitted in both tests	170	135	5	4
			36	
Total	3278	3766	100	100
Error of tabulation	2	—6		
Midscores	36-46	43-47		

For instance, approximately 36% of the responses in the Stanford Reading Test No. 2 fall in the region of variable control, in which effort and adjustment are factors determining score as well as ability. Therefore, scores in all such tests must be accounted mere indices of behavior and not measures of elemental factors.

Intelligence Tests. The Binet-Simon tests, and their many revisions, constitute performance scales in which to the natural difficulty of interpreting the index of behavior secured is added great possibilities of variation in conditions of giving and scoring. Finding a score in the Stanford revision of the Binet tests is a process of making a "summation of credits and penalties." The rule is⁶² "(1) Credit the subject with all the tests below the point

⁶²Terman, L. M. *The Measurement of Intelligence*. (1916). page 137.

where examination begins (remembering that the examination goes back until a year group has been found in which all the tests are passed); and (2) add to this basal credit, 2 months for each test passed successfully up to and including year X, 3 months for each test passed in XII, and 4 months for each test passed in XIV, 5 months for each success in 'average adult.' and 6 months for each success in 'superior adult.' "

The justification of intelligence testing is its utility. "The proof of the Binet method is in the fact that it works so well."⁶³ Intelligence tests are useful instruments for separating the individuals who differ in their gross intellectual endowment, but in no sense can they be said to "measure" intelligence. They measure instances of complex behavior of which some natural factor of the individual, called intelligence, is so large and so persistent a contributing factor that for purposes of gross discrimination more intelligent persons can be separated from the less intelligent. But the isolation and true measurement of this distinctive factor, the possession of which makes for intelligent behavior, seems to me to await the analytical treatment of measures of different levels of intelligent behavior by appropriate rate tests.

Illustration. When a series of tests, each of which differs widely in one thought element from the one before it, is given to a group of children, results are secured which at least seem to offer opportunities for analysis. For instance, a certain test in arithmetic⁶⁴ contained among others four related examples. The first example presented the elements of a situation in the sequence in which they are usually met; the second inverted the order in that one element and the sum is given to find the other element; the third expressed the same situation as a condition; while the fourth eliminated all except the abstract elements. The children did the work in their heads, merely drawing a line around the right answer among those given. (Figure 20.)

⁶³Terman, L. M. *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence*. (1917) page 153. Otis, A. "Some Logical Aspects of the Binet Scale," *The Psychological Review*, Vol. XXIII, No. 2, March; No. 3, May, 1916.

⁶⁴From an unpublished study by C. Louis Thiele, Supervisor of Exact Sciences, Detroit Public Schools.

FIGURE 20
Portion of Test Illustrating Differences in
Complexity in Test Elements

Detroit Public Schools		Exact Science Department			
MEASUREMENT OF EFFECTS OF COMPLEXITY OF THOUGHT PROCESS					
Type	Problem	Answers			
1. Direct					
	Jack went to the store. He bought a pencil for 5¢ and a tablet for 8¢. The tablet and the pencil cost _____ cents.	5¢	8¢	5¢	8¢
		3¢	5¢	8¢	3¢
		8¢	3¢	13¢	5¢
2. Inverse					
	Jack went to the store. He bought a tablet that cost 7¢. He gave the storekeeper 10¢. Jack received _____¢ change.	7¢	10¢	10¢	7¢
		10¢	3¢	7¢	3¢
		17¢	7¢	3¢	10¢
3. Conditional					
	Helen has 9¢. She wants to buy a doll. The doll costs 15¢. Helen must have _____¢ more to buy the doll.	15¢	15¢	24¢	15¢
		9¢	6¢	15¢	9¢
		24¢	9¢	9¢	6¢
4. Abstract					
	A tablet costs 10¢. A pencil costs 6¢. A tablet costs _____¢ more than a pencil.	10¢	6¢	10¢	10¢
		4¢	4¢	6¢	6¢
		6¢	10¢	4¢	16¢

In example 1, the elements are presented in the sequence in which they occur. In 2, the second element, 10¢ must be resolved into two. In 3, a condition is introduced. In 4, only abstract statements are made.

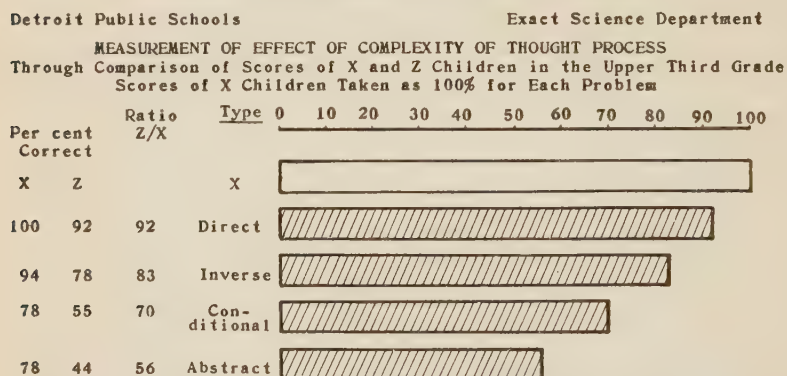
The children were asked to draw a line around the right answers.

The results show that the Z children, or the lower twenty per cent of the class, resemble the X children, or the upper twenty per cent, more closely (92%) in simple, direct behavior situations than in the situation involving understanding of abstract (56%) relations. (Figure 21.)

It is much easier to believe that intelligence contributes in a different degree to the two situations than it is to believe that children have a different type of intelligence for each type of problem. It is probable, therefore, that an analysis of each type of situation by rate tests in which the effect of various factors which are known to be functioning, such as motor ability, reading ability, etc., might be eliminated, would reveal a constant intelligence factor operating in all of them.

FIGURE 21

Comparison of Performance in Situations of Different Degrees of Complexity as an Index of Relative Intelligence



Children in the X group constitute approximately the upper twenty per cent of the class in intelligence and achievement; those in the Z group, the lower twenty per cent. Scores are the median per cents of accuracy of the two groups. In each case the ratio of the Z to the X score has been found; i. e. $44\% = 56\%$ of 78% .

As the complexity of the thought processes increases, the performance of the Z children differs more and more from that of the X children.

The solution of this problem, however, is merely a special case of Problem 5 already discussed, the determination of relative difficulty of rate tests of different degrees of complexity.

Reading. One further proof will be offered for variation in the results from performance scales, a comparison of the scores in performance tests and in rate tests for the subject of reading. (Table XXIX.) Selection of fifty pairs of equal scores was made from a large number of records of children in the Stanford total for reading. The scores of these same children in other reading tests were found and compared. They were then divided into two groups, A and B, one member of each pair being placed in each group.

Groups A and B thus have identical ability in reading, as determined by the Stanford reading tests. The question immediately arises, "What scores will these two groups have in the other reading tests?" These other tests are (1) another performance scale, the Thorndike-McCall reading test; (2) the

TABLE XXIX
Comparison of Variation in Scores in Rate Tests Made by Groups Matched in Stanford Reading Performance Scales

Pair Number	Scores in Tests									
	Stanford Reading Scales		Thorndike-McCall Reading Tests		Detroit Rate Tests		Courtis Reading Tests Kitten Series			
	A	B	A	B	A	B	Rate of Reading		Rate of Answering Questions	
							A	B	A	B
1	214	214	68	59	240	266	315	250	56	61
2	155	155	48	52	190	268	240	255	40	45
3	231	231	54	81	260	138	341	227	40	60
4	230	230	81	61	334	232	390	227	65	60
5	247	247	77	72	268	332	350	290	70	70
6	144	144	46	48	142	130	230	210	50	60
7	187	187	56	59	234	194	197	252	67	65
8	208	208	52	68	236	408	294	503	70	55
9	219	219	64	68	272	386	314	367	70	70
10	198	198	64	64	138	158	202	293	50	55
11	231	231	61	61	194	272	253	383	35	70
12	222	222	56	59	210	226	368	152	65	50
13	217	217	64	61	260	212	322	264	87.5	55
14	195	195	68	56	170	228	210	284	40	75
15	208	208	61	77	286	208	308	198	60	30
16	168	168	50	50	176	128	182	192	50	42
17	201	201	68	59	314	190	454	236	55	55
18	213	213	64	56	196	214	205	227	50	65
19	190	190	59	56	408	214	352	328	65	60
20	157	157	56	56	140	134	192	158	65	40

(Continued)

TABLE XXIX (Continued)

Pair Number	Scores in Tests									
	Stanford Reading Scales		Thorndike-McCall Reading Tests		Detroit Rate Tests		Courtis Reading Tests Kitten Series			
	A	B	A	B	A	B	Rate of Reading		Rate of Answering Questions	
							A	B	A	B
21	173	173	56	48	328	132	226	161	70	30
22	182	182	48	50	206	132	308	233	45	45
23	172	172	48	48	182	206	190	318	40	55
24	233	233	81	64	142	202	379	272	75	60
25	159	159	46	48	138	222	243	334	70	70
26	221	221	68	68	190	184	214	262	53	50
27	226	226	61	59	328	258	368	278	82.5	63
28	226	226	59	68	264	194	272	284	65	46
29	122	122	42	50	140	132	256	186	40	47
30	161	161	52	56	208	170	253	211.5	55	45
31	172	172	50	44	208	206	348	249	75	40
32	139	139	50	48	170	230	147	260	70	50
33	200	200	54	59	188	194	224	199	59	62
34	254	254	77	64	198	202	228	233.5	70	63
35	181	181	52	68	196	260	251	283	50	35
36	237	237	56	77	272	328	352	331	67.5	50
37	124	124	43	50	126	90	144	214	30	35
38	205	205	64	64	162	198	252	282	52	65
39	185	185	52	59	232	270	336	394	65	45
40	193	193	54	72	198	126	228.5	272	60	50

(Continued)

TABLE XXX
Comparison of Variation in Scores in Rate Tests Made by Groups Matched in Detroit Rate Tests

Pair Number	Scores in Tests											
	Stanford Reading Scales		Thorndike-McCall Reading Tests		Detroit Rate Tests		Courtis Readg Tests Kitten Series					
	C	D	C	D	C	D	Rate of Reading		Rate of Answering Questions		Index of Comprehension C D	
							C	D	C	D		
1	190	208	59	68	408	408	352	503	65	55	92	98
2	187	231	59	61	194	194	252	253	65	35	100	97
3	198	231	64	81	138	138	202	227	50	60	91	87
4	217	231	64	54	260	260	322	341	87.5	40	89	98
5	231	219	61	64	272	272	383	314	70	70	97	100
6	190	213	56	56	214	214	328	227	60	65	94	93
7	155	247	52	77	268	268	255	350	45	70	100	100
8	184	245	72	81	264	264	269	398	70	80	100	97
9	201	155	59	48	190	190	236	240	55	40	96	93
10	237	173	77	56	328	328	331	226	50	70	98	95
11	172	182	44	48	206	206	249	308	40	45	98	92
12	139	161	50	56	170	170	147	211.5	70	45	87	95
13	205	254	64	77	198	198	282	228	65	70	100	97
14	193	124	72	43	126	126	272	144	50	30	92	50
15	254	233	64	64	202	202	233.5	272	63	60	98	96
16	122	173	50	48	132	132	186	161	47	30	60	80
17	182	161	50	54	132	132	233	246	45	40	95	96
18	83	150	40	48	142	142	158	175	55	50	65	93
19	154	208	48	77	208	208	304	198	45	30	80	100
20	159	159	48	48	222	222	250	334	70	70	95	83

(Continued)

(Continued)

TABLE XXX (Continued)

Pair Number	Scores in Tests											
	Stanford Reading Scales				Thorndike-McCall Reading Tests				Detroit Rate Tests			
	C	D	C	D	C	D	C	D	C	D	C	D
Courtis Reading Tests Kitten Series												
										Rate of Answering Questions		Index of Comprehension
										C	D	
21	237	118	61	46	230	230	271	142	59	30	95	90
22	219	130	56	54	212	212	275	177	70	35	100	87
23	172	161	50	52	208	208	348	253	75	55	91	86
24	185	230	52	61	232	232	336	227	65	60	98	100
25	213	181	64	52	196	196	205	251	50	50	86	98
26	200	226	59	68	194	194	199	284	62	46	90	87
27	144	233	46	81	142	142	230	379	50	75	82	98
28	120	148	46	54	132	132	186	236	55	40	87	94
29	122	157	42	56	140	140	256	192	40	65	85	85
30	103	65	39	46	94	94	80	93	35	25	87	23
31	159	115	46	44	138	138	243	168	70	30	79	96
32	243	226	59	59	264	264	278	272	70	65	98	95
33	130	214	52	59	266	266	266	250	55	61	72	91
34	166	214	46	68	240	240	356	315	60	56	96	96
35	151	130	54	44	140	140	176	255	60	50	78	100
36	206	172	52	48	182	182	232	190	60	40	97	95
37	142	205	54	64	162	162	266	252	50	52	98	84
38	172	154	48	50	206	206	318	354	55	45	83	96
39	192	226	52	59	258	258	208	278	80	63	94	96
40	128	119	44	43	134	134	256	172	40	42	98	92

(Continued)

TABLE XXX (Concluded)

Pair Number	Scores in Tests									
	Stanford Reading Scales		Thorndike-McCall Reading Tests		Detroit Rate Tests		Rate of Reading		Rate of Answering Questions	
	C	D	C	D	C	D	C	D	C	D
	Index of Comprehension		Index of Comprehension		Index of Comprehension		Index of Comprehension		Index of Comprehension	
41	184	215	54	50	194	194	248	288	80	70
42	177	233	59	68	208	208	272.5	322	60	70
43	162	196	59	54	146	146	231	164	45	45
44	181	200	68	64	260	260	283	400	35	65
45	184	93	52	48	192	192	209	241	66	45
46	198	173	64	52	262	262	414	267	70	45
47	227	113	68	54	132	132	274	110	70	50
48	188	187	54	44	264	264	344	236	75	55
49	213	171	61	52	264	264	311	212	67	61
50	230	187	68	59	182	184	287	234	66	65

Detroit Rate test, a rate test so simple that it is believed ability may be assumed to be represented by the rate score alone without serious error, since practically all scores fall in the field of perfect control; and (3) the Courtis Standard Reading Test No. 2, Kitten Series, a simple reading test yielding a score in rate of reading, and also scores in rate of answering questions about what was read and in quality of answers (index of comprehension).⁶⁵

A second selection of fifty pairs was made, this time on the basis of equal scores in the Detroit Rate test.⁶⁶ The two groups formed from these pairs will be called Groups C and D.

The mean values for the various groups in the different tests were found. (Table XXXI.)

TABLE XXXI

Constancy of Scores in Performance Scales and in Rate Tests

Comparison of Mean Scores in Groups of 50 Made Equal in Various Reading Tests

Group	Performance Scales		Rate Tests			
	Stanford	Thorndike-McCall	Detroit	Kitten Rate	Kitten Question	Kitten Comprehension
A	189*	57.0	217	273	57.0	90.2
B	189*	58.1	208	262	54.6	91.5
C	180	55.6	203*	262	59.2	90.6
D	183	57.2	203*	251	52.2	91.6

*Tests in which groups were made equal.

The means of the rate tests differ more than the means of the performance scales, but only slightly more in most cases. However, a different story is told by the variabilities.⁶⁷

The rate tests are roughly twice as variable as the performance scales, but the groups made equal by performance scales are just about as variable as those made equal by rate tests. The lowest variability of all is in the comprehension scores in the rate test.

⁶⁵Illustrations of these tests will be found in the appendix.

⁶⁶Table XXX.

⁶⁷Variability is taken as standard deviation divided by the mean.

TABLE XXXII
Comparison of Variabilities of Rate and Performance Tests
 Results from same data as Table XXXI

Group	Performance Scales			Rate Tests		
	Stanford	Thorndike-McCall	Detroit	Kitten Rate	Kitten Question	Kitten Comprehension
A	.19*	.16	.28	.26	.23	.16
B	.19*	.15	.32	.24	.20	.10
C	.21	.16	.29*	.24	.20	.10
D	.24	.19	.29*	.31	.27	.14

*The tests in which the groups were matched.

These are just as constant for groups C and D, made equal by rate tests, as they are for groups A and B, made equal by performance scales. The explanation throws much light on the whole situation. Performance scales owe their apparent constancy of score, their so-called reliability, to the fact that they discard rate scores, make use only of the items right, and combine the results from many little tests into one resultant score. Their constancy is a statistical constancy, rather than a constancy of behavior. The method of scoring is so chosen that the scores are constant; but these relatively more constant scores do not measure ability in this field any more than quality alone measures ability in handwriting. The ideal test of the future will make use of both rate and quality in determining ability in performance scales as well as in rate tests.

Other significant facts about variation in performance scales are shown by correlations between the scores in the various groups. (Table XXXIII.)

The rate tests all yield low correlations, both between themselves and between the others. This is as it should be, since rate tests were devised to reveal every fluctuation in functioning of any of the factors involved. But that the correlations of the two groups selected to be equal in score in the Stanford tests, should be but .49 in the Thorndike tests, (Item 5) when the correlations between the two tests is .74 or better, can mean only that a large amount of variation is represented in these

TABLE XXXIII

Rate Tests and Performance Scales Compared as to Variability in Score by Means of Correlation

Groups A and B were made equal in Stanford Scores.

Groups C and D were made equal in Detroit Rate Scores.

Product moment correlation coefficients based on scores in Tables XXIX and XXX. Cases in each group = 50.

Comparison	Correlation	
	Coefficient	P. E.
1. Stanford Group A with Thorndike Group A.....	.76	.04
2. Stanford Group B with Thorndike Group B.....	.74	.04
3. Detroit Rate C with Thorndike Group C.....	.38	.08
4. Detroit Rate D with Thorndike Group D.....	.27	.09
5. Thorndike Group A with Thorndike Group B.....	.49	.07
6. Thorndike Group C with Thorndike Group D.....	.23	.09
7. Stanford Group A with Detroit Rate A.....	.44	.08
8. Stanford Group B with Detroit Rate B.....	.48	.07
9. Stanford Group C with Detroit Rate C.....	.45	.08
10. Stanford Group D with Detroit Rate D.....	.46	.07
11. Stanford Group C with Stanford Group D.....	.25	.09
12. Detroit Rate A with Detroit Rate B.....	.33	.08
13. Stanford Group A with Kitten Rate A.....	.48	.08
14. Stanford Group B with Kitten Rate B.....	.28	.09
15. Detroit Rate C with Kitten Rate C.....	.65	.06
16. Detroit Rate D with Kitten Rate D.....	.63	.05
17. Kitten Rate A with Kitten Rate B.....	.16	.09
18. Kitten Rate C with Kitten Rate D.....	.34	.07
19. Stanford Group A with Kitten Questions A.....	.36	.08
20. Stanford Group B with Kitten Questions B.....	.36	.08
21. Detroit Rate C with Kitten Questions C.....	.37	.08
22. Detroit Rate D with Kitten Questions D.....	.43	.08
23. Kitten Questions A with Kitten Questions B.....	.08	.09
24. Kitten Questions B with Kitten Questions D.....	.20	.09
25. Stanford Group A with Kitten Comprehension A.....	.94	.01
26. Stanford Group B with Kitten Comprehension B.....	.58	.06
27. Detroit Group C with Kitten Comprehension C.....	.32	.09
28. Detroit Group D with Kitten Comprehension D.....	.39	.08
29. Kitten Comprehension A with Kitten Comprehension B.....	.08	.09
30. Kitten Comprehension C with Kitten Comprehension D.....	.10	.09

scores in spite of the steps taken to prevent it. Also, it should be noted that the inter-correlations between the rate scores are uniformly higher between groups made equal on the basis of scores in the extremely variable rate tests than between groups made equal on the basis of scores in the so-called "reliable" performance scales (Items 12, 15, 16, 17, 18, 23, 24); while the correlation between the comprehension scores are practically the same for the two types (Items 29, 30).

Conclusions. In the light of all the evidence, it seems to me the judgment must be made that performance scales, useful as they are for certain purposes, are not suitable for the final determination of relationships between factors, and the ultimate science of education must await the solution of the problems of validity, utility, reliability, and difficulty in connection with both performance scales and rate tests. The imperfections of our present measuring instruments are, therefore, major causes operating to render difficult the discovery of exact relationship between factors in the educational field.

CHAPTER VIII

The Measurement of Heredity, Maturity and Training

The preceding chapter has presented data which prove that all tests measure directly only certain complex aspects of behavior. The purposes for which measurements in education are made, however, tempt us to go behind the face of the returns and draw inferences in regard to the effects of various types of factors. The existence of certain of these factors is admitted by all. No rational person denies that individuals differ at birth, that children change inevitably from birth to death, nor that individuals are affected by the experiences through which they are made to pass. The factors of heredity, maturity, and training are real, their effects are real, and the drawing of inferences in regard to them is proper. The question to be considered, therefore, is not "Do these factors condition a child's success in school?", but "To what degree do they condition his success?"

In view of the inadequacies which have been shown to exist in our present methods of evaluating performance, it should be evident that the attempt to carry further the analysis of existing measures of performance must, at the outset, be limited to gross approximations to the truth. The major purposes of the analyses discussed in this chapter are two: (1) to determine the precision of prediction in terms of existing measures of performance, and (2) to locate and define the problems which must be solved before the present degree of precision can be increased.

The terms heredity, maturity, and training have already been defined. At this point, however, it is necessary to discuss the criteria by which the nature of specific manifestations of these types of factors may be recognized.

Since the term heredity is to be applied only to that which is dependent upon the elements received from parents, it follows that permanence and constancy in manifestation are characteristics by which the hereditary elements may be recognized. Mere descriptive criteria yield less reliable inferences than direct observation and experimental manipulation under the law of the single variable, but even descriptive criteria may aid in making discriminations. Therefore, that which appears fixed and beyond control will be considered as due to heredity, although in the last analysis it must be admitted that the fixation may have taken place subsequent to the fertilization of the egg cell and prior to birth, and may be due in fact to environmental influences. Heredity thus becomes a general category in which are placed those persistent influences which do not appear subject to direct modification or control by man, and whose reactions to forces which are under control appear to determine the effect of those forces.

Training is a term applied to experiences which appear to be completely under control, and which are consciously manipulated as causes to produce desired effects. There is nothing inevitable about training. Its amount and character may be varied at will. Its effects are determined by heredity, maturity, and other factors, and so are beyond control, but the essential characteristic of training is that it itself may be manipulated or controlled.

Maturity is a term which is applied to forces which are intermediate between heredity and training, and partake of the nature of each. From one point of view, the process of maturation seems to have the inevitable character of heredity. Development appears to follow a fixed cycle or sequence in time. On the other hand, maturity is the result of the play of environmental forces upon hereditary elements and is conditioned by environment. To this degree it appears subject to control. Under ordinary conditions the environmental elements are so uniform that the uniformity of the process of development forces itself upon the attention. The term maturity, therefore, connotes in general, "present stage of development," having in mind both the initial and the ultimate condition of the organism.

The importance of the concept of maturity lies in the fact that behavior is a complex to which both physical and mental elements contribute. These may vary independently. If a person has a quick mind but a slow body, performance will be slow, since all performance involves movement. On the other hand, rate of movement may be very rapid but thought processes slow. In such cases, performance is again slow. Thus, two performances outwardly identical really have very different significance.

For instance, one of the series of tests given in this investigation was called the Motor-Mental Tests.¹ Test No. 2 of this series required of a subject merely that he copy figures; Test 3, that he copy letters. The tests were immediately repeated (as Tests 6 and 7), but on the second trial, a new thought element was introduced into each. In Test 6 (a repetition of Test 2), the children were asked to write, not the figure seen, but the figure which ordinarily follows that figure. Thus, if the figure to be copied was 4, the children were expected to write 5, 3 for 2, and so on. In Test 7 (repetition of Test 3), in place of writing the letter seen, the children were required to write the letter that ordinarily comes before that letter. Thus, if the letter to be copied was "g," the children were expected to write "f," and "a" for "b," and so on.

The results from three half-minute trials of the series of tests reveal a number of significant features. (Table XXXIV.)

TABLE XXXIV

Variations in Relative Maturity in Motor and Thought Processes

Individual	Midscores of Three Trials				Class Scores			
	Test 2 Copying Figures	Test 3 Copying Letters	Test 6 Writing Following Figure	Test 7 Writing Preceding Letter	Of Corresponding Level of Achievement in Test 2 and Test 3			
					Test 2	Test 3	Test 6	Test 7
A	110	120	64	32	} Above highest class average			
B	112	120	74	6				
C	94	90	50	12	}	94	83	53
D	94	92	52	24				
E	60	60	26	14	}	63	49	31
F	63	50	30	6				

¹See Appendix.

Individuals A and B, although approximately equal in the motor reactions of mere copying, are of very different ability in their mental reactions. A can think of the letter that comes before another letter comparatively readily, but B has difficulty in doing so. In similar fashion, C has more difficulty than D, and F than E. In describing such conditions, A, D, and E will be spoken of as more able in this function than B, C, and F.

Immediately the question arises as to what this superiority in ability is due. Is the difference between A and B caused by a difference in the mental machinery inherited by these two children? Is it caused by a difference in the degree of development of the inherited machinery, or by the degree of exercise (training) the machinery has had?

If the children are of the same age and have had the same practice (that is, if it is known that two of the variables are constant), then the difference will be ascribed to heredity. This would probably imply, also, the concept of different maximal developments, since the children have inherited different potentialities. Each child, therefore, may be equally mature in terms of its own inheritance. Thus maturity is a relative term.

If the children are of equal inheritance but of different ages, the inherited mechanisms will be considered to be at different stages of development; for even though both have received the same training, the training could not have had the effect on the undeveloped organism that it did on the more mature organism. The effect of training would seem to be limited by the development of the mechanism which receives it. In this case, the children would be considered to be of different ability, and the reason for these differences would be the differences in their physical maturities.

If the children are of equal heredity and equal maturity, then the differences would be considered due to differences in training.

In other words, in a situation in which a given result may possibly be the resultant of one or more of these variables, whenever two of them are known, differences will be ascribed

to the third. But in absence of knowledge of two of the factors, the situation cannot be analyzed with certainty.

Fortunately, experimental investigations have thrown some light on the process of maturation.² A number of freshly hatched chicks were divided into two groups. One group (C) was allowed to feed in the normal way, and each day the number of swallow-

TABLE XXXV

Maturation of the Pecking Instinct in Chicks

Average number of swallowing reactions in a daily series of twenty-five trials at pecking

Day	Control Group	Group B	Group D
2	3.7	1.0	
3	8.2	3.0	
4	10.0	5.8	
5	12.9	4.4	
6	12.8	5.8	
7	12.9	9.2	4.0*
8	12.8	8.6	9.3
9	15.4	10.2	13.0
10	14.9	9.0	15.7
11	15.4	6.4*	14.3
12	16.4	8.8	16.3
13	16.2	8.2	
14	16.1	12.0	
15	16.9	12.2	
16	17.1	13.8	
17	16.3	14.2	
18	18.0	17.8	
19	16.9	17.8	
20	16.7	16.3	
21	17.4		
22	17.0		
23	17.6		
24	16.6		
25	17.6		
26	18.1		

*End of forced feeding.

Control Group—normal feeding conditions.

Group B—Forcibly fed. Given twenty-five trials at pecking a day.

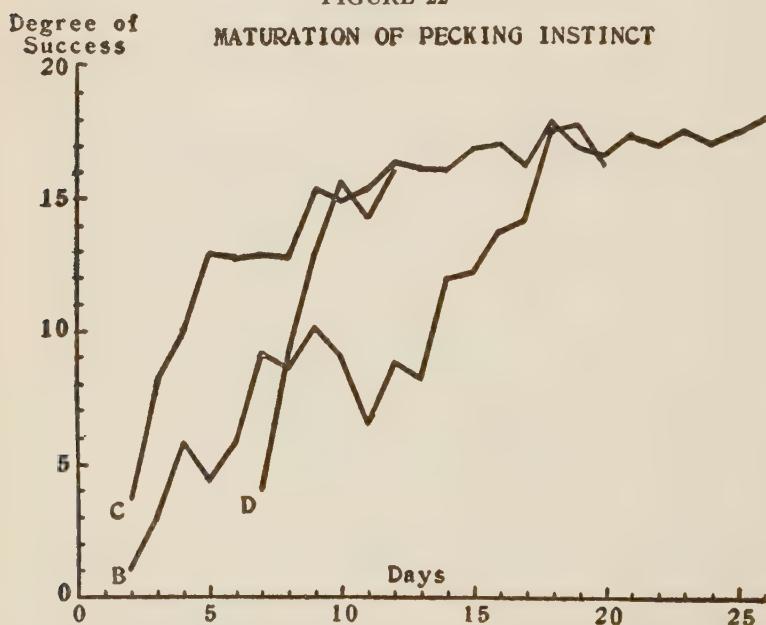
Group D—Forcibly fed for six days. No practice for six days. Normal conditions for the rest of the time.

²Bird, Charles. "The Relative Importance of Maturation and Habit in the Development of an Instinct." *Pedagogical Seminary*, (1925) Volume XXXII, Number 1. The reader is advised to study the author's own comments, as these do not entirely agree with the conclusions above.

ing reactions in twenty-five trials at pecking at grain was noted and taken as a measure of the maturity of the pecking instinct. The remaining chicks were kept in the dark and divided into a series of groups, only two of which will be discussed. Group B was forcibly fed, but once each day for ten days was given twenty-five trials at pecking. Group D was also forcibly fed and given no trials whatever for six days. It was then turned loose to eat under normal conditions. On the 11th day, Group B was also put under normal conditions. (Table XXXV.)

The control group shows a development curve of diminishing daily growth (Figure 22). Group B, which received only limited

FIGURE 22



Line C represents growth in skill of chicks under normal conditions. (Training is in excess of matured capacity. Consequently growth curve is of the usual form.)

Line B represents the growth in skill of chicks who received special practice of twenty-five trials a day (limited training) until day eleven, and normal practice thereafter.

Line D represents the growth of chicks who received no practice for six days and normal practice thereafter. Groups B and D were forcibly fed until restored to normal conditions.

Note the rapid growth until ability equals matured capacity.

practice, shows limited growth, but as soon as the amount of practice becomes normal, it begins to grow rapidly until it overtakes the control group. Group D is particularly illuminating. It had no practice for 6 days, and on the 7th, under normal practice, showed only as much ability as the control group at the end of the second day. In five additional days of normal practice, however, it overtook the control group which had taken 10 days to reach the same level of development. Moreover, Group B, as soon as the period of forcible feeding and limited practice is over, begins to grow rapidly, and in 9 days attains the level of development reached by the control group in 15 days. (From day 3, score 8.2, the score nearest to 6.4 the score of Group B at the beginning of normal practice to day 18, the first day when both groups have been approximately the same score.) This and similar experiments tend to show that training, acting on a developed or mature mechanism, produces larger effects than on an immature mechanism.

There is need, therefore, for the concept, matured capacity. Capacity, it will be remembered, was defined as potential ability. It is fixed by heredity and matured by experience. If now the organism has a number of experiences which develop the physical aspects of the mechanism, but which do not provide opportunity for the mechanism to function, actual ability to function will be nearly zero, but the possibilities of learning will be large. Consequently, training will produce at once large effects, much larger than the same training applied to a less developed mechanism.

Reference has already been made to the differences in the way teachers mark the older and the younger children. Children intellectually gifted often have an intellectual maturity which exceeds the powers of their immature physical mechanism to express it. They are, to this extent, underestimated. Their performances constantly fall short of their real intellectual maxima, because of the hampering effects of their inadequate physical mechanisms. On the other hand, the older dull children in a class often have physical mechanisms whose motor possibilities exceed their intellectual development. They tend to be

overrated by teachers. Their performances constantly reflect their intellectual maxima. It becomes important, therefore, to analyze performance in terms of the three concepts, capacity, matured capacity, and training.

An important aspect of both maturity and training is age. The passage of time is objectively measurable, and the entire organism, in all its parts, changes with time. When there is no specialized training, and the organism changes with time, it will be considered to be maturing. When time passes without appreciable developmental change, training being zero or in excess of needs, the organism will be said to be mature. At any time that continued training fails to yield growth, the individual will be thought of as trained to the limit of matured capacity. Plateaus in the learning curve may represent such conditions.

Where conditions of living are normal, developmental experiences are ordinarily in excess of matured capacity (as with the normal chicks) so that growth tendencies appear stable. Under such conditions, age becomes a measure of the matured capacity, and heredity the factor to which differences in the effect of age are ascribed. That is, under ordinary developmental conditions, training is uniform for all, since it is in excess of matured capacity, and differences in growth may be taken as due to heredity. In other words, under stabilized conditions, other things being equal, it is possible to predict outcomes in terms of age, but any shift in the stability of the conditions upsets all such predictions. Prediction, therefore, is to be thought of as made under a given set of conditions, and to hold only for those conditions.

An illustration may serve to make these relationships plain. Stuttgart, Germany, before and after the war, may be taken as representing stable conditions. Under such conditions, boys 11.5 years old in the Volksschulen tended to be 136 centimeters (53.5 inches) in height and to weigh 30.5 kilograms (67.1 pounds)³. (Table XXXVI, Figure 23.)

The height and weight of boys of various ages were affected to different degrees by the scarcity of food. The reasons for the

³*Grosse und Gewicht der Schulkinder und andere Grundlagen für die Ernährungsversorgung.* Herausgegeben vom Deutschen Zentrallausschuss für die Auslandshilfe. Berlin, (1924) page 45.

TABLE XXXVI

Variation of Height and Weight with Change in Conditions

Heights (centimeters) and weights (kilograms) of boys 11.5 to 12 years old in the Volksschulen, Stuttgart, Germany. Number of cases 600-700

Year.....	1913 1914	1915 1916	1916 1917	1917* 1918	1918 1919	1919 1920	1920 1921	1921 1922	1922 1923	1923 1924
Height.....	136	137	138	132	135	135	135	136	136	137
Weight.....	30.4	30.3	30.1	28.9	30.3	30.3	30.3	30.6	30.7	31.2

*Year of the Armistice.

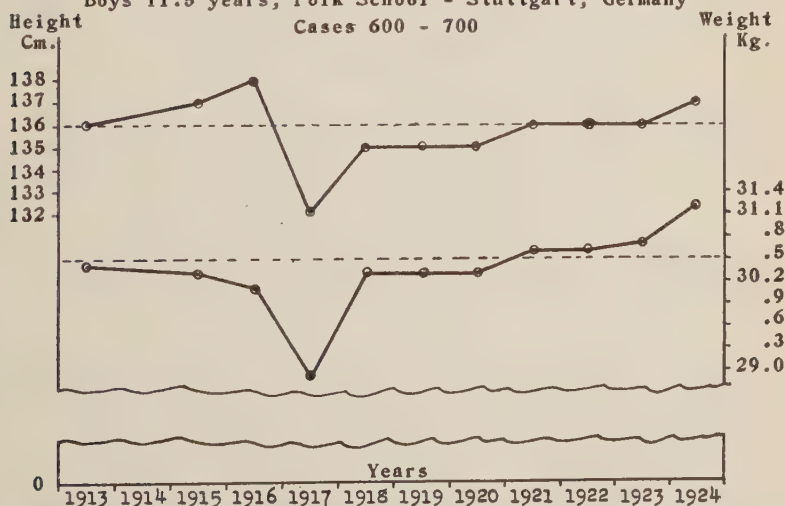
The data have been interpreted at their face value. It is extremely improbable, however, that the variations given in the table represent the true variations in height and weight, since they exceed the average yearly growth in amount. Selection caused by sickness and changed social conditions probably played a large part in determining the actual figures in this and the following table. Nevertheless, since other sources of information support the conclusion that the height, weight, and general development of school children were seriously affected by the food shortage, the figures are given as an illustration of a fact but not as a true measure of it.

FIGURE 23

Effect of Scarcity of Food upon Height and Weight of Boys,
11.5 to 12.0 Years of Age

EFFECT. OF WAR

Boys 11.5 years, Polk School - Stuttgart, Germany
Cases 600 - 700



For boys 11.5 years old, height and weight are fairly stable at 136 centimeters and 30.5 kilograms, but during the period of scarcity of food both height and weight show a large decrease; weight was affected more than height.

To prove that this fluctuation is not some chance effect, it is only necessary to compare the 1917-18 values for every half year with the values for the year 1923-24. (Table XXXVII, Figure 24.)

TABLE XXXVII

**Comparison of the Effects of Scarcity of Food
upon Children of Different Ages**

Heights (centimeters) and weights (kilograms) of Boys of various ages in the Volksschulen, Stuttgart, Germany

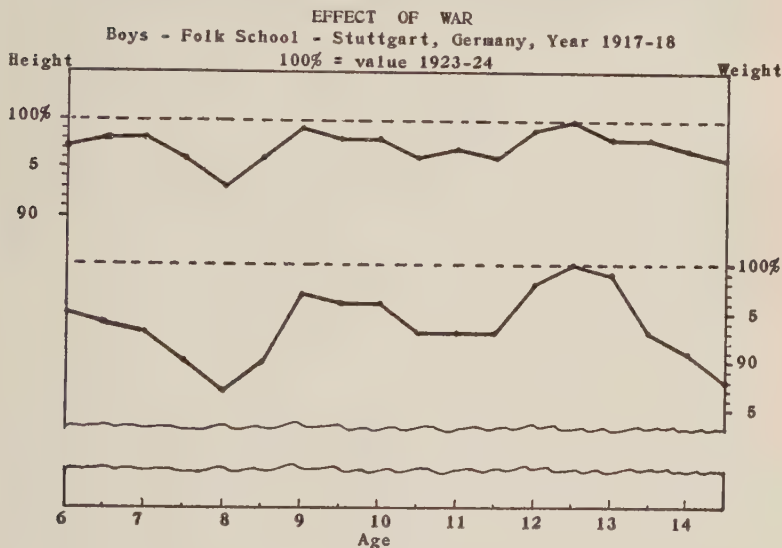
Age	Year 1917-18		Year 1923-24		Per Cent Year 1923-24 Taken As 100%	
	Height	Weight	Height	Weight	Height	Weight
14.5	145	35.6	151	40.5	96	88
14.0	143	34.9	147	38.5	97	91
13.5	143	34.7	146	37.1	98	94
13.0	142	34.5	144	34.9	99	99
12.5	141	33.4	141	33.5	100	100
12.0	138	31.7	139	32.3	99	98
11.5	132	28.9	137	31.2	96	93
11.0	131	27.8	135	29.8	97	93
10.5	128	26.9	133	28.8	96	93
10.0	128	26.5	131	27.5	98	96
9.5	127	25.6	129	26.6	98	96
9.0	126	25.3	127	26.1	99	97
8.5	119	22.8	124	25.2	96	90
8.0	115	21.0	123	24.1	93	87
7.5	116	20.9	121	23.1	96	90
7.0	114	20.4	119	21.8	96	94
6.5	113	19.6	116	20.8	97	94
6.0	111	19.4	114	20.3	97	96

variation are not known. For some ages, the effect is greater the following year. Weight was generally affected more than height.

Height in any stable situation is regarded as determined by heredity, but only in the sense that variations from average height are due to hereditary factors. When the stability of the situation is broken up, as by shortage of food, the average height, and individual variation from the average, change correspondingly. The part played by environment in determining height is thus made evident. If the new conditions endured and became stable, height at maturity would have new values.

FIGURE 24

Comparison of the Effects of Scarcity of Food upon
Heights and Weights of Children of Different Ages



The illustration should serve to make plain the difficulties encountered when measurement of any phase of an educational situation is attempted. Each phase influences other phases, and all are so interdependent that it is impossible to determine the effect of any factor singly.

Suppose, for example, one attempts to measure the effect of training upon score in intelligence tests. Score is affected by sex, age, intelligence, and other factors as well. If the attempt is made to hold these factors constant by studying the scores of boys, of the same age, and of approximately the same intelligence, and social status, but differing in amount of training received, only a very few such cases can be secured and only for a limited difference in training. (Table XXXVIII.)

The results reveal no tendencies that can with certainty be ascribed to a single variable. All that can be said is that if training does produce any effect upon scores in intelligence tests,

TABLE XXXVIII

Determination of the Effect of Training upon Intelligence Scores

Score in National Intelligence Test of Selected Groups of Boys and Girls of the Same Ages, but Differing in the Number of Semesters Spent in School. Total Cases 258

Age	Training one unit less than normal				Training Normal			Training one unit more than normal			
	Cases	Units	Score	Difference from Normal	Cases	Units	Score	Cases	Units	Score	Difference from Normal
13.0	30	14	131	-7	30	15	138	6	16	140	+2
12.5	5	13	120	-13	56	14	133	5	15	153	+20
12.0	27	12	117	-16	28	13	133	7	14	143	+10
11.5	14	11	106	-12	45	12	118	5	13	135	+17
Mean difference from normal				-12							+12

The effect of one semester of training for years 11.5-13.0 appears to produce a difference in intelligence score of 12 units or approximately 9% of the mean normal score for these ages. If, however, the scores were tabulated on the basis of grades attained, the children in the higher grades would have the higher scores, because the school by promotion tends to sort the bright from the dull. In the table, the children have had different numbers of semesters of training because, among other causes, of differences in the age at entrance which in itself may have caused the differences. Many such factors are known to be operating. No certain conclusions can be drawn from such data.

the effect is compounded with the effects caused by the other factors whose operation within the various groups has made the selection possible.

Clearly what is needed is a statistical process by which a more perfect analysis may be made. The processes in common use are partial and multiple correlation.

The law of the single variable is absolute. Whenever an effect *may be* the result of more than one factor, it is not possible to determine the effect of any one of the factors with certainty until it has been proved that the factor is the sole cause of the effect. Statistical methods themselves operate as causes affecting results. Scientific caution demands, therefore, that the effects of variations in statistical assumptions and treatment be determined. Hence there emerges

Problem 7. Validity of Statistical Procedures

Nature of the Problem. Mathematical analyses are developed according to logical principles. Certain assumptions are made and formulae deduced. If the principles of logic have been adhered to, the resulting formulae are perfectly valid statistically.

When, however, the attempt is made to put these formulae to use, a difficulty arises. Unless the situation conforms to the assumptions under which the formula was deduced, the formula does not fit. Unfortunately, it is perfectly possible to make computations in accordance with general rules, but using data from situations which do not conform to assumptions made. In such cases, the results obtained do not represent the truth. Validity applied to a statistical method is to mean "appropriate to the situation" in which it is used.

Illustration. For instance, the Pearson product-moment method of computing a correlation coefficient rests on the assumptions of rectilinearity of relationship, of homogeneity of data, and of reliability of individual measures. If it is used in situations in which the data are not homogeneous, the relationship between the variables curvilinear, and the measures unreliable through errors, the results while correct mathematically misrepresent the true relationship to the degree that the situation departs from the assumptions upon which the mathematical processes are based.

To check the validity of certain correlation formulae, I obtained from a high school class in chemistry the results from an experiment in the determination of the equivalent of magnesium. The experiment involved four variables, weight of magnesium, observed volume of gas set free when magnesium was put in acid, the temperature, and the pressure. The weight of the magnesium was converted into "true volume of gas," and the attempt made to determine by the correlation technique the relationship between the factors.⁴ The total number of cases was 29. Correlation coefficients were obtained as follows:

⁴These are known. They are called Boyle's Law and Charles' Law. The first is that volume is inversely proportional to the pressure; the second, that volume varies directly with temperature.

TABLE XXXIX

Relation Between True Volume, Observed Volume, Temperature, and Pressure

Correlation Coefficients

	True Volume	Observed Volume	Temperature	Pressure
True Volume.....	..	.16	.23	.15
Observed Volume.....	.16	..	.02	.03
Temperature.....	.23	.02	..	— .26
Pressure.....	.15	.03	— .26	..

If now the attempt is made to determine the relation between pressure and observed volume, holding temperature constant, or between temperature and observed volume, holding pressure constant, by the partial correlation technique, coefficients are obtained differing but slightly from those in the table, in spite of the fact that such correlations under the laws known to be operating should be -1.00 and $+1.00$ respectively.

The explanation is, of course, that the relationship between pressure and volume is curvilinear, and the values obtained by a class of beginning students are unreliable, so that the mathematical procedures used are not suited to the situation. The true formula⁵ for the relationship between these factors is

$$Y = C_1 \frac{X_1 X_3}{(C_2 + X_2)} \quad \text{in which}$$

Y = True volume

X_3 = Pressure

X_1 = Observed volume

C_1, C_2 = Constants

X_2 = Temperature

If nothing were known except the observed volumes of the gases, and the attempt were made to predict the corresponding true volumes, the best value to give would be the average value of the true volume, 164.13 cc. This means that in attempting to predict the true volume from the observed volume alone, the

⁵The formula is a mathematical expression for predicting true volume from the known relations as formulated in Boyle's and Charles' laws. It ignores one factor, vapor tension.

standard error of prediction would be the standard deviation of the true volumes, 28.57 cc., or an error of 17.4%.

If now prediction is made for all the factors by the formula

$$T.V. = C_0 + C_1 Ob. V + C_2 T + C_3 P$$

in which T.V. = true volume, Ob. V = observed volume, C_0 , C_1 , C_2 , C_3 = constants, T = temperature, P = pressure, the standard error of estimate would be decreased only to 26.87 cc. On the other hand, if perfect⁶ data are used with this same formula, the predictions are perfect to the fourth decimal place.

Need. There is need, therefore, for the constant application of objective checks upon statistical analyses to determine how valid are the statistical procedures employed. At present, the correlation technique yields mass results indicative of general tendencies. These are of great value, but the divergent observations are lost in the general average, and the procedure itself affords no direct opportunity to determine which observations depart from the general rule, and why.

Solution. I believe progress will be made only through the substitution in statistical analysis of analytical methods for our present mass or average methods. As has been shown, a given objective score which measures a complex may be produced by the combination of the different factors in many different relative proportions. There is need, therefore, to study the changes which take place in individual observations as they are corrected for one factor at a time.

Prediction Formulae. To illustrate the type of statistical analysis I have in mind, I shall present the derivation of a prediction formula in several different ways. I shall use as an illustration the prediction of weight of boys from age and height, and limit the scope of the prediction to ages 10 to 13 inclusive. The complete Weight-Height relationship is curvilinear; weight increases relatively more rapidly than height with age. By limiting the ranges of ages, the relation line may

⁶That is, observed volumes computed to exactly correspond to true volumes modified by the given pressures and temperatures.

be taken as rectilinear within the range without serious error. The basic data are the heights (to the nearest whole inch), the ages (to the nearest month), and the weights (to the nearest quarter-pound), of 190 boys, 10 to 13.

The mean weight of the group is **77.12** pounds, the mean age **142.7** months, and the mean height **56.36** inches.

The standard deviation of the weights is **13.02** pounds.

If nothing were known about a child except that he was a boy between the ages of 10.0 and 13.5, the best prediction that could be made from the data of this investigation would be to estimate his weight as the average weight of such boys, or 77.1 lbs. The prediction formula is

$$Y = 77.1 \dots \dots \dots (A)$$

On this basis, the standard error of estimate would be 13.02 lbs.

If the age of the boy was given, a better prediction could be made. The correlation between age and weight for the 190 cases is given by the coefficient r , $= .61 \pm .03$. The prediction formula is given by the equation:⁷

$$Y = C_0 + C_1 A \dots \dots \dots (B)$$

or, substituting the values found for the constants,

$$\text{Weight} = -5.70 + .60 \text{ Age}$$

The weight of each of the 190 boys was predicted by this formula, and the standard error of these estimates was 10.93 lbs.

Weight, however, is determined by height as well as age, and if the age and height of a boy are given, a still closer prediction can be made. The prediction formula is

$$Y = C_0 + C_1 A + C_2 H \dots \dots \dots (C)$$

or $\text{Weight} = -114.56 + .03 \text{ Age} + 3.33 \text{ Height}$

Prediction by this formula reduces the standard error of estimate to 6.89 lbs.

⁷The values of the constants were determined by the method described by Huntington, E. V., "Curve-Fitting by the Method of Least Squares and the Method of Moments," in *Handbook of Mathematical Statistics*, (1924), page 67, edited by H. L. Rietz.

In other words, by making use of a larger number of factors, the standard error of estimate can be reduced from 13.02 to 6.89, a reduction of 47%.

The average boy 144 months (12 years) old is 58 inches tall⁸ and weighs 85 lbs. By formula (A) his weight would be 77.1 lbs., an error of 7.9 lbs. By formula (B) his weight would be 80.7 lbs., an error of 4.3 lbs. By formula (C) his weight would be 82.9 lbs., an error of 2.1 lbs.

These results illustrate clearly how with increasing knowledge of the factors acting prediction can be made which is closer and closer to the truth. The size of the standard error of estimate is a measure of how adequate the knowledge is upon which prediction is based. Perfect knowledge would enable one to make perfect predictions. The standard error of prediction would then be 0.

It follows that since the standard error of estimate by formula C is reduced only 47% of initial error, there are other factors affecting Weight which must be known before perfect prediction can be made. The multiple correlation coefficient for weight with age and height is but .87. Yet there are no other general factors to be considered. Therefore, the remaining factors must be individual.

The conventional correlation method used thus far is expressed in *average values* representing general tendencies. What is needed is an individual, particularized method which will enable the investigator to follow the changes in each value as correction is made for one factor after another. The proportionate effect of each factor can then be determined, and by a study of divergent cases, the nature of such factors as may be operating on individuals discovered.

For instance, the piecemeal development of such a correction formula for weight is as follows:

The first prediction formula is the same as A above, namely

⁸Emerson, Wm. R. P., and Manny, F. A. "The New Weight-Height Tables and Malnutrition." *Archives of Pediatrics*, (1924) XLI:10, October.

$$Y = 77.1 \text{ lbs.} \dots\dots\dots (A)$$

or, Weight = 77.1.

This means that the average of weight for the boys of this group is 77.1 lbs. But weight is not developed in isolation. As a child matures, age, height, and other factors change as well. The full statement which recognizes the fact of dependent variation is somewhat as follows:

An average weight of 77.1 is the concomitant of an average age of 142.7 months and an average height of 56.36 inches and of a similar average value for each factor that may be considered, since all are but specialized aspects of a *single development*, boy. In other words, if a boy is 142.7 months old, his weight, on the average, will be 77.1 lbs. If it is not, some special factor must have caused the variation. Or to look at the same fact from a different point of view, if age varies from 142.7 months, weight will vary correspondingly. Therefore, the average age, 142.7 months, was subtracted from each age, the amount and sign of the deviations being found. In similar fashion, the average weight was subtracted from each particular weight, the amount and sign of these deviations being recorded. Next, these two sets of deviations were compared as to relative magnitude and relationship. In this case, upon inspection it was judged that the deviations in weights were, on the average, simple multiples of the deviations in heights. The value of this constant was found by the formula

$$C = \frac{XY}{X^2}$$

in which C equals the average value of the constant desired, X one variable and Y the other. In this case, the value of C was found to be .602. Therefore, formula A above becomes

$$Y = 77.1 + .602 (A - 142.7) \dots\dots\dots B^1$$

or, in words, weight equals the average weight + .602 times the difference between the age and the average age. This formula reduces to

$$Y = -8.8 + .602 \text{ Age} \dots\dots\dots B^{11}$$

equivalent of B above.⁹

If each weight is predicted from each age and the error of prediction found, a second order series of deviations will be obtained, which will be known as the deviations of actual weight from "weight predicted from age." These represent the parts of the deviations of weight from average weight which are not accounted for by the deviations of age from average age. They must be referred to other factors.

If now we attempt to study the deviations of height from average height as the cause of unexplained deviations in weight, we shall be comparing height deviations from which the effect of age has not been removed. For height matures with age as well as weight. Accordingly, the next step is to correct the height deviations for age in the same manner as the weight deviations were corrected. The formula proved to be

$$H = 56.36 + .159 (A - 142.7) \dots\dots\dots B^{11}$$

or, in words, height = 56.36 inches plus .159 times the deviation of age from average age.

Each height was predicted by this formula and the second order deviations for height found. These represent the part of the height deviations not due to age. Since, however, these differences between individuals are not known to be caused by any general factor other than age, they will be called measures of the hereditary elements in height until such time as they can be further resolved. That is, a boy who comes from parental stock which is taller than average will be taller than average at every age, and a boy who comes from stock which is shorter than average will be shorter at every age. (Figure 25.)

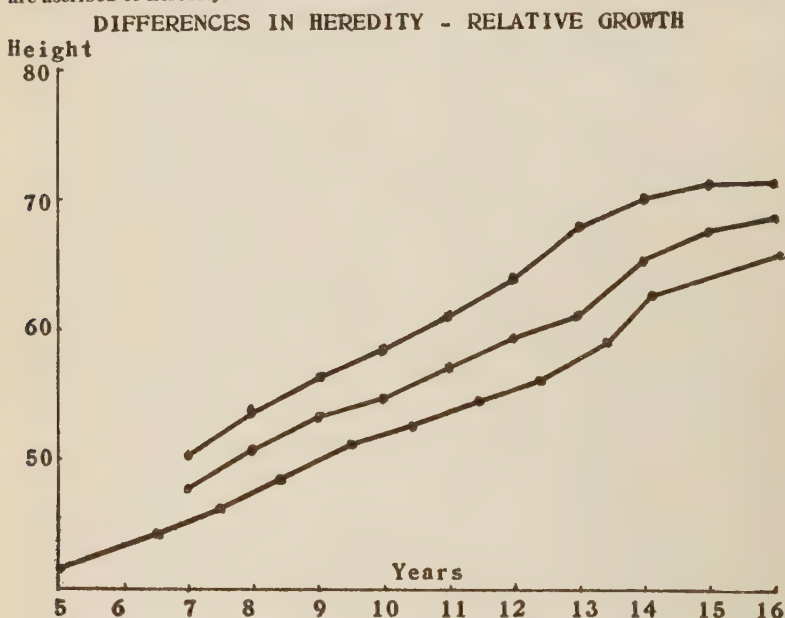
Comparing the second order of age deviations in weight with the corresponding "ageless" deviations in height, the average relationship is found to be

$$W = 3.34 H \dots\dots\dots (C^1)$$

⁹The differences between formula B and B¹¹ are caused by the differences in the number of situations in which figures are dropped. The sum of these errors amounts to a total of approximately 3% in the case of the boy under consideration.

FIGURE 25
Effect of Heredity upon Height¹⁰

Development curves of a tall, average, and short boy. Relative differences in height, being present at birth, persisting throughout life, and not being subject to control are ascribed to heredity.



or, the weight deviations not caused by age are on the average equal to 3.34 times the height deviations not caused by age. Prediction by this formula yields a series of third order deviations which need to be explained. The amount and frequency of these deviations are as follows:

TABLE XL
Errors in Weights Predicted from Age and Height

	—16 and less	—12	—8	—4	0	4	8	12	16 and over
Amount.....									
Frequency....	5	12	32	47	56	19	9	5	5

Total Cases, 190

Median Value, —.09

¹⁰Baldwin, B. T. "The Physical Growth of Children from Birth to Maturity." *University of Iowa Studies*, (1921) Volume I, No. 1, pages 73, 96. The data for the curves above are selected from those given by Baldwin, and redrawn, the values being scaled from his graphs. Unfortunately, Baldwin does not give the measurements directly.

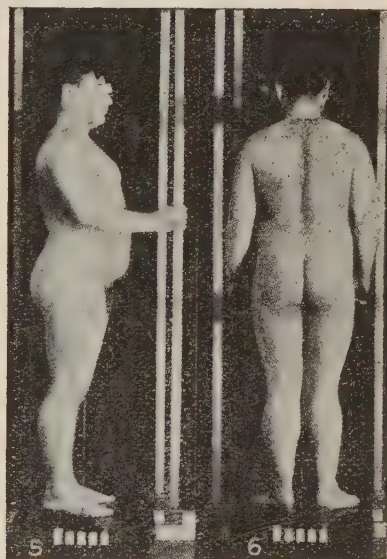
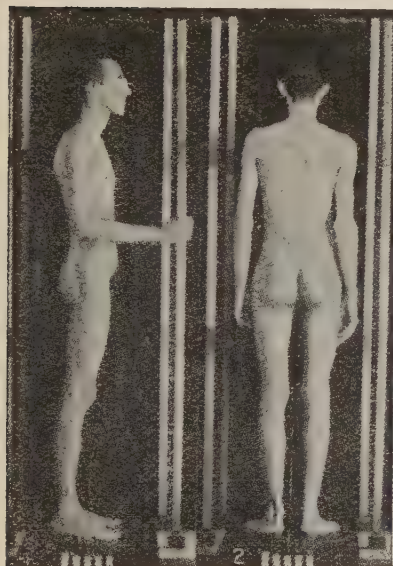
Two commonly recognized causes of individual variation in weight from normal are (1) build and (2) condition of health.

The concept of build has been discussed by many writers.¹¹ The most recent and comprehensive discussion of build presents a scale, of which the two extremes are shown here.¹²

FIGURE 26
Extremes of Scale for Build
(Davenport)

Tall — Thin

Short — Stocky



Davenport presents evidence to prove that differences in build are present at birth¹² and persist throughout life. The measure of build he proposes for children, chest girth $\div$ by stature, I was unable to use because measurement of chest girth was not made in this investigation. Furthermore, the measure itself is not constant at all ages or even for the restricted period under consideration. It is probable, therefore, that the index suggested is merely an index, not a measure: that it represents a

¹¹Baldwin, B. T. "The Physical Growth of Children from Birth to Maturity." *University of Iowa Studies*, (1921) Volume 1, No. 1, pages 207, 218-19.

¹²Davenport, C. B. *Body Build and Its Inheritance* (1923) Plates I and II.

complex, of which build may be the major factor. The measure given as second best, weight $\div$ by stature,¹³ was found for 190 cases. Its use will be explained a little later.

The second factor causing variation in weight is the condition of health. Everyone knows that, within limits, weight fluctuates according to supply of food, healthiness, etc. However, since no measure of healthfulness was secured, no direct use could be made of this factor.

For the purpose of this illustration, all the third order weight deviations were assumed to be due to build, and an index of build computed by the formula

$$B = W_{h-a}/W_{h-a} \dots\dots\dots (D)$$

which is to be read: Build is equal to the ratio of the actual deviations of weight corrected for the effects of height and age to the weight deviations predicted from the height deviations corrected for age. This measure represents a hereditary factor peculiar to each individual. Normal build is represented by 1.00; the actual values obtained were

TABLE XLI
Variation in Computed Build Factor

Value of Index	-6 and less	-4	-2	0	+2	+4	+6 and over	Total	Med- ian
Frequency	5	5	28	119	24	2	5	188 ¹⁴	+ .94

Formula B¹ above amended by the inclusion of C¹ and D becomes

$$Y = 77.1 + .602 (A - 142.7) + 3.34B[(H - 56.36) - .159 (A - 142.7)] \dots (E)$$

¹³Ibid. Page 14.

¹⁴In two cases there was no weight deviation from the prediction from age and height.

$$= - 8.8 + .602 A \\ + 3.34B[H - .159 A - 33.67] \dots\dots\dots (E^1)$$

or if build is excluded

$$Y = - 121.3 + .069 A + 3.34H \dots\dots\dots (E^2)^{15}$$

the equivalent of formula C above.

Formula E is to be read as follows: The weight of any boy between the ages of 10 and 13.5 is equal to 77.1 lbs. + .602 times the difference between his age and 142.7, (the average age,) plus 3.34 times the build times the quantity, height minus the mean height, (56.36 inches,) minus .159 times the age minus the mean age, (142.7.)

The standard error of estimate of prediction by means of this formula is 0 for the weights, heights, and ages from which it was obtained.

For instance, a certain boy was 126 months old, 53.5 inches tall, and his build factor was 2.34. By substituting in the formula his weight is found to be 65.55. His actual weight was 65.5. Error .1 of 1%.

The value of a formula, however, consists not in its effectiveness in predicting known, but unknown results. Accordingly, for each of the 190 boys the age and height in 1924 and the build factor from 1923 were used to compute the weight in 1924. The standard error of these estimates was 10.58 lbs., or 55% of the standard deviation of the weights (19.26). The standard error of prediction without the build factor was 7.97 lbs., or 41% of the standard deviation.

The explanation is not difficult. The build factor is not the true build factor but an assumed one.

In the attempt to secure a better measure of build, the weight deviations from age and height were correlated with "weight divided by stature" and "weight divided by stature square." The results gave negligible correlations, $r = .03$ and $.01$ respectively, so that neither of these measures of build is adequate

¹⁵The weight of the average boy 12 years old predicted by this formula is 82.4 lbs.; predicted by formula C, 82.9 lbs., a difference of but .6%. In other words, as the number of factors increases the difference caused by the differences in method of derivation tend to disappear. It can be shown that the two methods are mathematically equivalent. See appendix.

to explain the errors of prediction of weight from age and height. However, there **is** a build factor, and the purpose of the present discussion is to show that analytical treatment reveals the need for it, and stimulates its discovery. Its ultimate discovery and use must await more extended investigations in this field.

For the 1923 results, the standard error of estimate by the best formula was 0, and for 1924, 7.97 or 14%.¹⁶ This means that when the weight, height, and age of a boy are known, and his height a year later is given, his weight can be computed only within 8 lbs. As more information is secured in regard to what characteristics are indicative of the hereditary factors, it should be possible to predict both height and weight under stable conditions from the measures of a few significant factors.

Criticisms. The discussion has been carried forward without comment in order to indicate the general method of piecemeal correction for factors. I now wish to criticise the procedure and indicate the problems which must be solved before it can be perfected.

The first defect is that caused by selection. Any small group represents but a small part of the total population. For a formula to be valid to predict the weight of any child of any age, the data from which it was derived would have to be truly representative. Actually, the ages were restricted, there were irregularities in the number for the different ages, and within each age there were irregularities in the distribution of the different heights and weights. As the constants in the formulae are average values to which all values contribute, an undue number of large values at one extreme or the other would change the constants correspondingly.

For instance, the average value of the relationship in formula B¹¹¹ above is given as .159. If the ratio is actually computed for each of the 190 pairs, and the ratios distributed, the distribution makes plain the irregularities which exist from faulty selection. There are 13 cases, 6 at one extreme and 7 at the other, which

¹⁶If 58 inches is considered the average height of twelve year old boys.

depart widely from the median value .198. The distribution as a whole is skewed slightly toward the lower values.

TABLE XLII

Variation in Ratios of Age Deviation from Average Age to Height Deviations from Average Height

Value of Ratio.	—2.0 and less	—1.5	—1.0	—.5	0	.5	1.0	1.5	2.0 and over
Frequency . . .	7	3	4	39	106	16	9	0	6
Total Cases, 190		Median Value, .198.							

In such cases, the median .198 is probably a closer approximation of the truth than the mean, .159, but the use of the median leads to large immediate errors of prediction. These may be the best values for this order of deviations. If so, their validity would be proved when the residual errors of the next order were correlated with build. However, lack of a valid measure of build prevents this check being applied. Therefore the mean value of the ratio was used in the formula. There are many different values which might be found for each constant, depending upon the assumptions made in regard to the proper selection of cases. Therefore, under Problem 7, the discovery of methods of validating such assumptions must be regarded as a basic sub-problem.¹⁷

A second sub-problem is presented by the mingling of the constant and variable factors representing maturity and heredity. For instance, a true measure of build would be a constant from birth to death. While the actual gross proportions of the body change as a child develops, such shifts may be explained as the result of environmental forces acting on the hereditary elements. In determining the relation between age and weight, the average value obtained is distorted by the presence of the constant build factor. Height itself is a complex, and there are probably build, or corresponding hereditary factors for each growing unit; as legs, trunk, etc. If a true measure of build were available, the first step in the correlation procedure

¹⁷An illustration of the use of the piecemeal derivation procedure in the detection of factors will be found in the appendix.

would be to make the weight deviations comparable by correcting each one for its own peculiar hereditary element. The relation between the corrected weights would be quite different from that between the uncorrected weights.¹⁸ Unfortunately in massed data the presence of these constant factors is often unsuspected until revealed by the residual errors of prediction after allowance has been made for every known factor. At present, therefore, the development of prediction formula can be only a process of trial and error. However, piecemeal development, with intensive study of divergent cases, offers a program that leads at least to the isolation of the fundamental problems.

Uses. The value of such prediction formulae merits some discussion. There is in America today much emphasis upon health education, and much is made of the contributions which healthy bodies make to intellectual progress. On the other hand, a disinterested survey of the studies in which attempts have been made to measure the relation between physical and mental development, yields negative results.¹⁹ (Table XLIII.)

TABLE XLIII

**Correlation of Physical Traits with Educational Achievement
Fourth Grade—Gates**

Ossification	= .16	Lung Capacity	= .00
Height	= .01	Grip	= .15
Weight	= .07	Nutrition	= .15
Chest Girth	= .03	Physical Vigor	= .38
		(Estimated)	

r, educational achievement with mental age	.595
R, educational achievement with (mental age + ossified ratio + height + weight + chest girth + lung capacity + grip + nutrition)	.626
Reduction of Standard Error of Estimate	
By use of M. A. alone	= .20
By use of all factors	= .22

¹⁸It may be well to state here the general principles which apply to all attempts to determine relations between the factors in any situation. First, separate the factors operating into two classes, general, or those which affect every case, and specific, or those which affect only particular cases. Second, correct all raw scores for the effect of specific factors. (This means, of course, that the laws for specific factors must be first determined in situations in which the general factors are constant.) Third, determine the relationship between the general factors from the corrected scores. Failure to correct for the effects of specific factors (attenuation) lowers the correlation and obscures the relationships. Correction of correlation coefficients for attenuation is, from this point of view, of little value. It is the raw scores that must be corrected before true relationships can be determined.

¹⁹Gates, A. I. "The Nature and Educational Significance of Physical Status and Mental, Physiological, Social, and Emotional Maturity." *Journal of Educational Psychology*, (1924) Vol. XV, Number 6, page 346.

"The Educational Significance of Physical Status and Physiological, Mental, Emotional, Social Maturity." *Teachers College Record*, (1924) Volume XXV, Number 3, Page 223. May.

The probable reason is easily discovered. None of the measures used are pure measures of health. The contributions of age, height, build, etc., represented in weight would probably obscure the effects of changes in health, unless the other factors were first removed. At the present time, there are no measures of health which have been validated.

On the other hand, children are being selected for nutrition classes on the basis of the average relationship of weight to height. Emerson says:

"The relationship of height and weight to age has a definite value, but the matter of primary importance is the relationship between weight and height. . . . The boundaries which we use for what we call the safety weight zone—twenty per cent overweight for height and seven per cent under weight—have been determined by clinical evidence. Their adequacy is still questioned at times, but we can only repeat here what we have often stated, that in our experience we have never found a single child who is habitually seven per cent under weight who does not manifest other unmistakable signs of malnutrition."²⁰

Such statements ignore the build factor. A study of the heights and weights of individuals through several years, show that many children have persistent relations between height and weight which differ markedly from those given in the usual tables, but which, nevertheless, are normal for them.

For instance, the actual weight record of one individual for five semesters is shown in Figure 27 by the heavy line, marked actual. The weights given for his height in a table constructed for the use of teachers in schools is shown by the light broken line, while his weight, computed by Formula C, is shown by the heavy broken line. Judged by these standards, the individual is about 30 pounds over weight. If allowance is made for build,²¹ however, his "predicted weight line" is close to his actual line. The shaded portion in the figure represents the deviation of actual from predicted weight allowing for build. At 11.0 years, the individual may have been slightly under weight

²⁰Emerson, Wm. R. R. and Manny, F. A. "The New Weight-Height Tables and Malnutrition." *Archives of Pediatrics*, XLI: 10, October (1924).

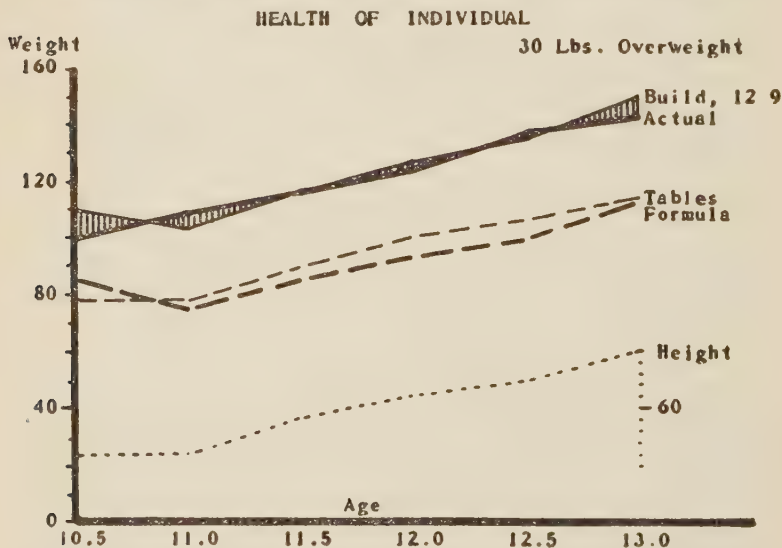
²¹Derived from the residual error of the 1923 predictions.

and in ill health. It is much more probable, however, that even this small variation is caused by imperfections in measurement, as height was recorded only to the nearest inch.

Variation of actual from predicted weight may be due, also, to variation in rate of growth at different periods. A careful determination of the exact general laws of growth would enable teachers of health education to study the effect of disease and other influences upon individual variation from predicted values. The desired measures of health may ultimately be found through the use of such prediction formulae.

FIGURE 27

Illustration of Use of Prediction Formulae in Study of Individuals



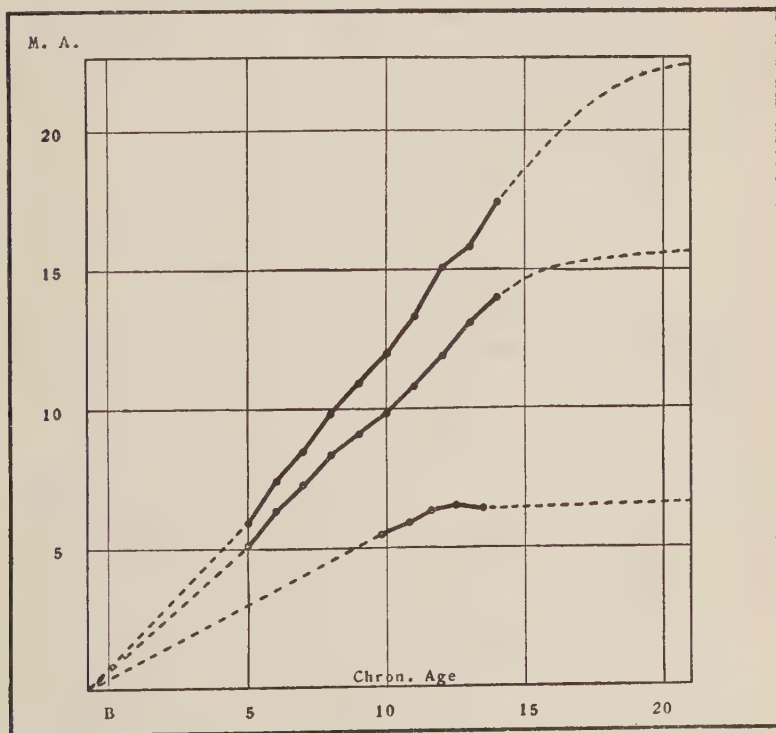
Line marked "actual" represents the weights of a boy very tall for his age and very heavy. His record for growth in height is indicated by the dotted line. By formula, without allowance for build, his predicted weight would agree fairly closely with the "weight for his height" given in a certain table prepared for the use of teachers in school, which makes him 30 pounds overweight. If a build factor is included in the prediction formula, the predicted weights are in substantial agreement with his actual weights.

It is suggested that variation from weights predicted by an "individual" formula that truly represents the general growth tendency (indicated by the shaded portions of the figure) will furnish a better index of healthfulness than variation from the average values given in the conventional tables.

Emerson's statement above is a significant acknowledgment of the fact that the validity of weight-height relationship rests on a basis of subjective opinion. The existence of cases like that shown in Figure 27 not only proves the need for objective determination of validity, but raises the question as to whether

FIGURE 28

**Hereditary Differences in Maturation of Intelligence.
Development Curves Based on Mental Ages**



The full development curves for intelligence as expressed in Mental Age (M. A.) are not known, and the evidence at hand is not consistent owing to changes in certain determinations of mental age, but it is possible to choose records which suggest the form of curves shown in the figure. The lower and upper portions of the curves (dotted), therefore, may be very different from those shown but sufficient data are at hand to warrant the statement that, in general, the curves as drawn reflect average conditions as we know them.

Redrawn from Baldwin and Stecher, "Mental Growth Curves of Normal and Superior Children: University of Iowa Studies (1922) Vol. II, No. 1.

or not all the conclusions so far reached in regard to the physical-mental relationship might not be overturned if truly valid measures were used in determining the relationship.

For one thing, scores in intelligence tests are measures of complexes. Such scores should probably be resolved into elements one of which would represent a hereditary factor analogous to build. The concepts Intelligence Quotient and Mental Age will at once come to mind. Mental Age Curves²² (Figure 28) exhibit all the characteristics of height curves. Something to which the name intelligence is given appears to mature with age, and the various somethings which contribute to score have constant relations with each other. Heredity and maturity are thus both represented in intelligence scores, which are thus mere measures of performance in intelligence tests. Under such circumstances, it is not surprising that little trace can be found statistically of a relationship which many "common sense" experiences would lead one to suppose exists.

One further use of prediction formulae will be indicated. A question frequently asked but seldom answered is "What is the effect of this or that factor?" Prediction formulae enable the answer to be found.

For instance, suppose the question is asked, "What is the effect of maturity upon weight from ages 10.5 to 13.0?" This question means how much of the growth from 10.5 to 13.0 is caused by age, build being normal (or 1.00) and height being normal and changing normally with age.

The average height for a 10.5 year old (126 mo.) is 55 inches; for a 13.0 year old 60 inches. His weights would be:

Weight
at Year

Age

$$10.5 = 77.1 + .602 (126 - 142.7)$$

Height

$$+ 3.34 \left[(55 - 56.36) - .159 (126 - 142.7) \right]$$

²²Baldwin, B. T. and Stecher, L. I. "Mental Growth Curves of Normal and Superior Children." *University of Iowa Studies*, (1922) Vol. II, No. 1.

Age

$$13.0 = 77.1 + .602 (156 - 142.7)$$

Height

$$+ 3.34 \left[(60 - 56.36) - .159 (156 - 142.7) \right]$$

or

Weight at

Year $13.0 = 90.2 = 77.1 + 8.0 + 5.1$

Weight at

Year $10.5 = 71.3 = 77.1 - 10.1 + 4.3$

Difference

$18.9 = 18.1 + .8$

In other words, when a boy of normal build increases 18.9 pounds in weight from age 10.5 to age 13.0, 18.1 pounds or 96% of the growth is due to mere change in maturity, and .8 pounds or 4% to factors represented in height in addition to the effect of the change in height due to maturity alone.²³

Prediction of Scores in the Stanford Tests. If from a field where all the factors are known and reliably measurable by standard scales having true zero points (physical magnitudes), we turn to a field, education, where the factors are only crudely measurable at present, the question naturally arises, "How far is it possible to determine what makes children succeed, and how much of the total growth should be attributed to each factor?"

A prediction formula for success²⁴ in school has been worked out by the piecemeal method for 165 boys between the ages of 10.5 and 13.5 based on heredity, maturity, and training, assuming rectilinearity of relationship. The formula is:

²³The reader will do well not to be deceived by this statement and by other similar statements to be made later on. If correction had been made for height first, then age, the relative percentages attributed to these two factors would have been reversed. Both age and height are functions of "general development" and in the statement as given, this factor is grouped with age. The 4% attributed to height is thus 4% over and above the general development factor which might have as truly been grouped with height as with age. The time will undoubtedly come when an adequate method of expressing the effects of general and specific factors will be invented. For the present, statements like that above tell a suggestive story if correctly interpreted.

²⁴Success being understood to mean, as defined, score in Stanford Achievement Tests.

$$\begin{aligned}
S = & 56.6 \\
& + .38(A - 144.2) \\
& + .50[(N - 119) - .58(A - 144.2)] \\
& + .40[(D - 61.0) - .42(A - 144.2)] \\
& \quad - .584[(N - 118) - .58(A - 144.2)] \\
& + .847[(T - 11.63) - .16(A - 142.7)] \\
& \quad - .016[(N - 118) - .58(A - 144.2)] \quad (F)
\end{aligned}$$

In which S = Stanford Score, A = Age in months, N = National Intelligence Scores, D = Detroit Army Intelligence Score, and T = Number of semesters of attendance at school.

Formula F condenses to

$$S = 2.43 - .07A + .253N + .40D + .847T \dots\dots\dots (G)$$

The standard error of estimate in predicting the scores of the 165 boys by this formula is 5.46, or approximately 10% of the average score. The standard deviation of the group of 165 Stanford scores was 13.24. The formula, therefore, reduces the standard error of estimate 59% and corresponds to a multiple correlation coefficient of, $R = .91$.

In other words, knowing a child's age, his scores in the National and Detroit Army Intelligence tests, and how many semesters the child has been in school, it is possible to predict his score with an average error of but approximately 4.4 points.²⁵ The average Stanford score of the group was 56.6 points. A child's success in school, therefore, can be predicted from the given data with an average error of but 8%. The score of three-fourths of the children can be predicted within 15%.

It will be remembered that the standard deviation of the changes in the Stanford scores on a retest after an interval of

²⁵Average deviation taken as .80 SD.

TABLE XLIV

Accuracy of Prediction of Stanford Scores

Predictions based on intelligence, age and semesters in school, or upon heredity, maturity and training.

Part A

Distribution of differences between predicted Stanford scores and actual Stanford scores for the 1923 tests from which the formula was derived. 165 cases.

Error.....	-16	-14	-12	-10	-8	-6	-4	-2	0	+2	+4	+6	+8	+10	+12
Frequency....	1	1	3	6	14	18	18	24	24	23	9	10	8	5	1
% of Total...	1	1	2	4	8	11	11	14	14	14	5	6	5	3	1

Half within 4 points, three-fourths within 5.5 points, nine-tenths within 9.2 points.

Part B

Distribution of same differences expressed as per cents of total score.

Error.....	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	+5	+10	+15	+20	+25
Frequency	1	0	1	0	1	4	4	21	23	30	31	21	22	3	3	
% of Total	1	0	1	0	1	2	2	13	14	18	18	13	13	2	2	

76% of the cases within $\pm 15\%$.

Part C

Distribution of differences between predicted Stanford scores and actual Stanford scores for the 1924 tests from which the formulae were not derived. Errors expressed as per cents. 69 cases.

Error.....	-25	-20	-15	-10	-5	0	+5	+10	+15	+20	+25
Frequency	1	3	6	11	11	14	11	6	2	3	1
% of Total	1	4	9	16	16	21	16	9	3	4	1

87% of the cases within $\pm 15\%$.

two days was a little under 7%.²⁶ This suggests that all the remaining factors which influence a child's score have a combined **general** effect which is not more than 3% of the total score.

The question of the relative contributions of the different factors may be answered as before. Computation of the score of two boys, one 13.0 years old making a National Intelligence Test score of 122, a Detroit Army score of 66.4, after 14 semesters of school training; the other 10.5 years old, National score 95.4, Detroit Army score 56.8, semesters of training 9 (data represent average values for the different years), gave results of 60.73 and 48.08 respectively, a gain in score of 12.65 points. This represents the average growth for this group during a two and a half

²⁶See page 101.

year period. Of this growth 11.4 or 90% is due to maturity (or the changes in total development brought about by the distinctive effects of age plus the general development factor) 1.2 or 9.5% to those aspects of intelligence which are not determined by change in maturity, and .5% to those aspects of training which are not determined by change in maturity and intelligence.

This then is the general answer to the problem of the dissertation. Children²⁷ succeed in school work in general in accordance with their development or maturity. The level of their development at any age level is fixed by hereditary factors which are measured roughly by the intelligence tests, while the training contributes a small but distinctive amount in addition. These factors account for 90% of the changes in children. Other factors, such as home influence, health, etc., cause variations in individuals, but on the average all such factors combined do not displace the average child more than 10% from his predicted place.

The attempt was made to trace the distinctive effect of such factors as attendance at Sunday school, social environment, etc., but no results were secured at variance with the conclusions above. Attendance at Sunday school is a "specific" factor which may produce enormous changes in one individual, but as a "general" factor operating on all individuals who attend, the effect is within the 3 or 4% limit suggested above. The same is true of the effect of differences in social environments.

The real test of a prediction formula is in the prediction of future results. The Stanford scores for as many of the boys as were present in the 1924 test²⁸ was predicted from the 1924 data for age, intelligence, and training. The standard error of the prediction was 6.8.²⁹ This result is small enough to justify both the conclusion that the factor concept is applicable to educational products and the belief that experimental attempts to solve the problems which have been pointed out are desirable.

²⁷Strictly speaking, boys between the ages of 10.0 and 13.5 in the schools tested. While specific formulae have not been developed for girls, or for children in other schools, enough comparisons have been made to warrant the assumption that the results obtained apply in a general way to "children."

²⁸Complete records were secured for 69 of the 165 boys.

²⁹A reduction of 51.5% in the initial variability, corresponding to a multiple correlation coefficient of $R = .87$.

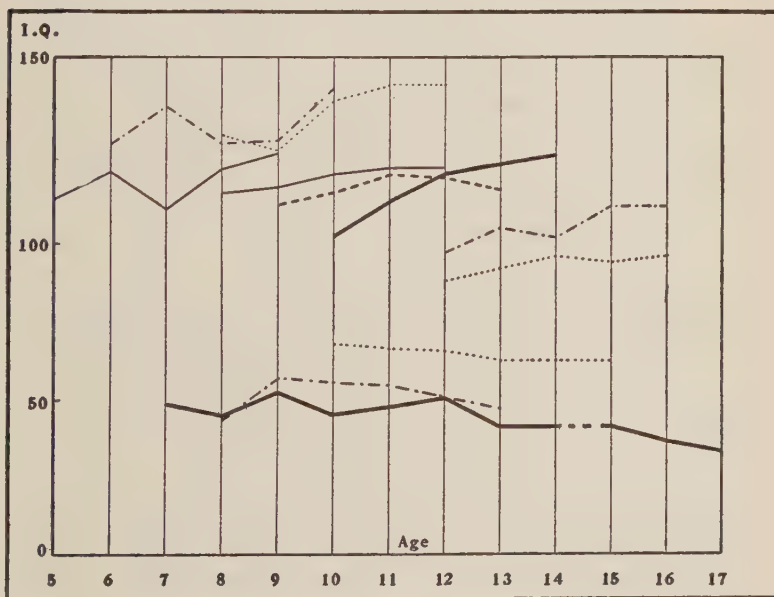
Criticisms

While predictions by the formula are surprisingly close, considering the many defects in the measuring instruments discussed in Chapter VII, no one will believe that present precision represents ultimate precision. Indeed, all the evidence leads to the conclusion that it will be possible shortly to greatly increase the certainty of prediction.

For instance, the best present measure of the constant hereditary intellectual factor is the Intelligence Quotient. As pointed out previously, the determination of the I. Q. is by a procedure which is open to many criticisms of unreliability. The studies of the reliability of the I. Q. by various investigators give high values

FIGURE 29

Variation in Successive Determination of I. Q's of Children



Data in the main selected from records given by Baldwin and Stecher, "Mental Growth Curves of Normal and Superior Children," University of Iowa Studies (1922), Vol. II, No. 1. One or two Detroit records included. Cases chosen to illustrate three types of variation: (1) steady increase in I. Q.; (2) irregular variation; (3) steady decline.

for massed data.³⁰ But records of repeated testing show considerable variation, and apparently of three sorts: (1) steady increases, due possibly to handicaps of language or similar factors in initial tests; (2) irregularities, probably due to variations in testing conditions; and (3) steady decreases, considered by some to represent normal conditions.³¹ (Figure 29.)

It was judged, however, that this situation offers no new problems beyond those already discussed. Accordingly, no attempt was made to introduce the intelligence quotient into the formula. It is probable that intelligence scores should first be made comparable by correction for specific factors before age and training relationships are determined.

In similar fashion, the emotional and social factors undoubtedly produce marked effects upon success. The residual errors of prediction of Stanford scores were correlated with the results from the emotional and social surveys but the correlations were practically zero. In my judgment, this means, not that these factors do not influence a child's success in school, but that they have no general or uniform influence. Proof of their effects is to be derived from analytical study of individual variations from predicted values.³² Methods of correction for such specific factors must be discovered before the true relations between the general factors can be determined.

Prediction of Scores in the Handwriting Tests. The Stanford tests are batteries of performance scales, which, while dealing with conventional school subjects, reading, arithmetic, spelling, etc., are similar in many ways to intelligence tests. Accordingly a prediction formula for handwriting based on midpoint scores for rate and quality reduced to comparable ability scores ($R + Q$) was developed. It is

³⁰Miller, W. S. "The Variation of Significance of Intelligence Quotients obtained from Group Tests." *Journal of Educational Psychology*, XV:6, September (1924).

Herring, J. P. "Avery's Comparison of the Stanford and Herring Revisions." *Journal of Educational Psychology*, XV:6, September (1924).

³¹Hart, H. "The Slowing Up of Growth in Mental-test Ability." *School and Society*, XX:573, November (1924).

Dvorak, A. "The Relation of I. Q. to the Prognosis of Special Class Pupils." *School and Society*, XIX: 736, June (1924).

Madsen, I. N. "Some Results with the Stanford Revision of the Binet-Simon Tests." *School and Society*, XIX:559, May (1924).

³²Note that in Table XLIV, plus and minus errors of 20% and more point to the existence of unknown specific factors.

$$\text{Hdw} = 107.0$$

$$+ 1.02(A - 142.7)$$

$$+ .289[(N - 118) - .58 (A - 144.2)]$$

$$+ .208[(M - 146.8) - 1.64 (A - 142.7)] \dots\dots\dots (G)$$

In which Hdw = ability in handwriting (rate plus quality), A = Age, N = Score in National Intelligence Test, M = Sum of rate of copying figures and rate of copying letters in Tests 2 and 3 of the Motor-Mental tests. Training was not used as when age was removed the correlation between training and the residual errors fell to zero. This probably means that in general the opportunities for training were uniform for all by being in excess of matured capacity.

Written in the conventional form, formula G becomes:

$$\text{Hdw} = -30.3 + .51A + .289N + .208 M \dots\dots\dots (H)$$

The standard error of estimate of the predictions made by the formula for the 1923 data is 18.0, corresponding to a reduction of but 20% in the initial error. The total change in ability from 10.5 years to 13.0 years was 31.7, of which 96% is due to maturity, 9% to intelligence, — 5% to motor ability.

The distribution of the errors of prediction of the handwriting scores (Table XLV) shows that 68% of the 122 cases of the 1923 results and 68% of the 49 cases of the 1924 results fall within a range of 15%. I interpret this to mean that the more specific the function measured, the more the relationship between the general factors is obscured by the effect of the specific factors. This accounts also for the wide range of errors (from — 75% to + 25%). Accurate prediction of scores in narrow specific functions must await improved technique in the analysis of the effects of general and specific factors.

An illustration of what is meant by specific factors is easily found. The fields of variation of different individuals center at different portions of the rate-quality area. That is, two indi-

TABLE XLV
Accuracy of Prediction of Handwriting Scores

Predictions based on Intelligence, Age, Semesters in School, and Rate of Motor Reaction.
Distribution of differences (expressed as per cents of actual scores) between predicted handwriting scores and actual handwriting scores for the 1923 tests from which the formula was derived, and for as many of the same individuals as were present in 1924.

1923 Results																	68%		
Error.....	-60 to	-51	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	15	20	25	Total
Frequency.....	4		0	1	0	2	2	4	3	20	15	14	5	18	11	12	6	5	122
% of Total.....	3		0	1	0	1	2	3	3	17	12	11	4	15	9	10	5	4	100
% of cases within $\pm 15\%$																			
1924 Results																	68%		
Error.....	-75 to	-51	-50	-45	-40	-35	-30	-25	-20	-15	-10	-5	0	5	10	15	20	25	Total
Frequency.....	1		1	0	2	0	0	4	1	3	5	4	8	8	5	3	3	1	49
% of Total.....	2		2	0	4	0	0	8	2	6	10	8	17	17	10	6	6	2	100
% of cases within $\pm 15\%$																			

viduals having the same measure of ability may have a very different relation between their rate and quality scores. Thus, Individual A may have a mid-point of 80 rate and 40 quality, resulting in an ability score of 120. Individual B, on the other hand, may have scores of 50 rate and 70 quality, which yields the same ability score, 120. Yet their centers are widely separated.³³ That such characteristics persist is shown by finding a measure of position for midpoint scores. This may be taken as the ratio of the rate of the actual mid-point score to the rate that is equal to the quality score. The formula is:

$$P = R / \frac{R+Q}{2} \text{ or } \frac{2R}{R+Q} \dots\dots\dots (I)$$

In which P stands for the position ratio of the mid-point, R = rate, Q = quality. That is, the position of mid-point is shown by the ratio of twice the rate to the sum of the rate and quality. Correlation of P, 1923 with P, 1924 for 117 boys gave $r = .40 \pm .05$. P thus represents a crude measure of a specific constant (hereditary?) factor.

Relation Between Hereditary Factors

An interesting study is suggested by the isolation of the various hereditary factors. Are they relatively constant; that is, does

$$\frac{H_I}{H_{Gp}} = \frac{B_I}{B_{Gp}} = \frac{I. Q_I}{I. Q_{Gp}} = \frac{P_I}{P_{Gp}} = C$$

when H = the hereditary constant in height, B = in build, I. Q. = in intelligence quotient, P = in position of mid-point handwriting score, I = Individual, and Gp = group?

In other words, if specific hereditary factors could be freed from the influence of all other factors, would they for any one individual tend to approximate a constant quantity, irrespective of the scores from which they were derived? If so, the result might be regarded as proof of Spearman's general and specific factor theory. The present study does not furnish even an

³³Difference = $\sqrt{2 \times 30^2} = 42.4$ units.

approximate measure of any of these quantities, but the possibility of such an investigation is almost worthy of a place among the fundamental problems raised by this dissertation.

Conclusions

The outstanding conclusions to be drawn from the data presented in this chapter are, in my judgment, (1) that persistent and consistent influences are operative in children's behavior, influences which while only crudely measured are nevertheless sufficiently well measured to enable prediction to be made with a greater degree of certainty than most people believe is possible; (2) that enough evidence of imperfection and uncontrolled variation in results has been shown to warrant the belief that very great improvements in prediction are possible if problems are attacked from the analytical rather than from the massed data viewpoint, and (3) that prediction formulae offer new methods of attack upon the problem of tracing the effect of such factors as specialized training in control experiments, etc. Taken all in all, the data suggest great possibilities for discovery and progress in the science of measuring educational products.

CHAPTER IX

Summary and Generalizations

The investigation undertaken to find out why children succeed in school work has yielded several different kinds of products.

A. Terms and Definitions.

The analytical treatment to which data have been submitted has made discrimination necessary and led to definitions and much conceptional thinking. The terms, Heredity, Matured Capacity, Maturity, Ability, Training, Effort, Adjustment, Conditions, Performance, have been given either new or more precise connotations.

B. The Factor Concept.

Analysis of records of behavior has been carried further and by different methods than in the past. The existence of tendencies sufficiently unitary and general to serve as bases of prediction has been shown, and the relations between our present crude measures of these factors determined within the limits of accuracy of the data.

C. Why Children Succeed.

Boys within the age range and school conditions studied have been proved to succeed in their school work to different degrees primarily because of differences in the maturity or development factor best represented by age. Scores in intelligence tests as measures of a general hereditary factor, serve as an important additional base of prediction, while training and other factors also have some small distinctive influences apart from the general developmental process. Approximately 7% of the present errors of prediction of Stanford scores have been

proved to be caused by imperfections in the measuring instruments themselves. The small portion remaining unexplained (3%) represents the combined distinctive effect of all other factors, such as teaching ability, social status, emotional control, home influence, etc. The total general effect of such factors is surprisingly small in amount, and furnishes no measure of their specific effects upon individual cases.¹

D. Problems.

Out of the many difficulties encountered have come the formulation of seven distinct problems which in effect map out a program for future work, a life program of research, so intricate and difficult that progress is likely to be made but slowly.

The seven problems are:

1. Validity or the determination of the characteristic nature of factors.
2. Utility, or the determination of the relative value of factors.
3. Reliability, or the control of variation in testing.
4. Quality, or the validation of measurements of quality.
5. Difficulty, or the evaluation of different levels of performance in common terms.
6. Effort, or the discovery or invention of methods of measuring directly the effort expended by the subject taking a test.
7. Validation of Statistical Procedure, or discovery of methods of proof of suitability of means to ends.

E. Methods.

Some progress has been made in modifying existing statistical methods and adjusting them to the situations peculiar to research in this field. In particular, the piecemeal development procedure for prediction formulae should facilitate the study of divergent cases and the location of the causes of divergence.

¹It must be remembered that the 10% residual error corresponds to 41% of the original variability.

Implications. From one point of view there is danger that too much has been proved. Analytical treatment that reveals the operation of law in the mental and spiritual realms often has a destructive effect upon idealistic thinking. Successful prediction sometimes gives rise to ideas of predestination or materialism, neither of which is conducive to joyful creative growth.

On the other hand, knowledge of law really operates in precisely the opposite direction. Knowledge, properly directed and controlled is power, and power and control mean freedom. Research reveals what is, not what may be. Therefore, the pessimist, he who permits fatalistic despair to inhibit effort, ought to take no encouragement from the results of this study. For the evil conditions which are revealed by research always belong wholly to the past. The future is yet to be made. When man shall have solved all his problems and, in the power of perfect knowledge, shall have attained to perfect control, he will be able to bend all elements to his will. The results of this study seem to me to mean that at present we are so tied and bound by ignorance that the world sweeps forward to a destiny fixed by natural tendencies uncontrolled by human effort. The road is strewn with failures and destruction. Some children achieve phenomenal success but, in the main, the efforts of both children and teachers move on to the inevitable "mechanical" result. The implications of this study are not that such mechanistic results are inevitable but that these conditions may be changed; that success and happiness for all await merely the solutions of problems, difficult and puzzling to be sure, but not insoluble. Approximately 25% of the children show evidence of being affected by specific factors to such a degree that their scores fall markedly above or below the predictable result. Knowledge of the factors causing these variations would make for control, and by intelligent adjustment of means to ends all of the children might be modified at will. But at present, for any stable situation, the success of children is predictable in terms of natural forces. Therefore I shall close by quoting the reflections of a master mind as he saw clearly, for the first time in human history, the picture which this dissertation has attempted to portray in more detail.

Nature and Nurture²

“There is no escape from the conclusion that nature prevails enormously over nurture **when the differences of nurture do not exceed what is commonly to be found among persons of the same rank of society and in the same country.** My fear is that my evidence may seem to prove too much, and to be discredited on that account, as it appears contrary to all experience that nurture should go for so little. But experience is often fallacious in ascribing great effects to trifling circumstances. Many a person has amused himself with throwing bits of sticks into a tiny brook and watching their progress, how they are arrested first by one chance obstacle, then by another; and again how their onward progress is facilitated by a combination of circumstances. He might ascribe much importance to each of these events, and think how largely the destiny of the sticks had been governed by a series of trifling accidents. Nevertheless, all the sticks succeed in passing down the current, and in the long run, they travel at nearly the same rate. So it is with life in respect to the several accidents which seem to have had a great effect upon our careers. The one element that varies in different individuals, but is constant in each of them, is the natural tendency; it corresponds to the current in the stream, and **inevitably asserts itself.**”

²Galton, Sir Francis. *Inquiries into Human Faculties*, (1883) page 131. The italics are mine.

APPENDIX

I	REASONS FOR THE CHOICE OF TESTS SELECTED.....	207
II	SAMPLE TESTS AND RECORDS.....	211
III	DETAILED INSTRUCTIONS FOR GIVING AND SCORING CERTAIN TESTS	231
IV	DETERMINATION OF FORMULAE FOR EFFORT AND ADJUSTMENT FACTORS.....	243
V	EQUIVALENCE OF PREDICTION FORMULAE.....	247
VI	ILLUSTRATIVE COMPARISON OF THE CONVENTIONAL AND PIECE- MEAL METHOD OF DERIVATION OF PREDICTION FORMULAE	251

I

REASONS FOR THE CHOICE OF TESTS SELECTED

There are apparently five types of factors which contribute to the success of a child in school: (1) Physical, (2) Mental, (3) Educational, (4) Emotional, (5) Social.

The first three are simpler in their nature and easier to measure than the last two. On the other hand, there are some indications that the last two are much more influential in determining success in certain situations than the first three. It was decided to limit the investigation to a careful measurement of the motor, mental, and educational abilities and to attempt experimental measurement of the emotional and social factors. Certain additional tests were given to secure special supplementary data.

The series of physical tests used was selected after conference with the Detroit Health Education Department as being those most likely to indicate the relative physical condition of the children, as well as best suited to reveal the effect of various factors upon performance.

Measurements of height, weight, lung capacity, and grip are relatively simple processes. For each there are standards, objective units, and a standardized procedure, so that the errors of measurement are relatively small. Height and weight are not performances in the ordinary sense, but the result of heredity, maturity, and conditions of development. The errors of measurement are almost wholly errors of observation and of conditions. Measurements of lung capacity and grip, however, while as simple and objective in nature as measurements of height and weight, are true measurements of performance, and are affected by the factors conditioning performance. Comparisons of the different types of results secured, therefore, serve to bring to light the effect of the introduction of extraneous factors upon the scores obtained.

Height and weight differ in still another respect. Height is almost wholly a matter of heredity and maturity. Fluctuations in height do not often occur. A child may increase in height irregularly, but changes take place slowly and once made are retained. Weight, on the other hand, fluctuates with the health of the individual. At one time he may gain in weight and at another lose. Weight thus varies much as performance varies. Analysis and measurement of the height-weight relationship, therefore, was undertaken to throw light upon similar variations in performance.

Measurement of performance in physical stunts, such as throwing a baseball for distance, is measurement of more complex performances. Physical, mental, emotional, and control elements enter into ball throwing, but the physical aspects of height, condition, etc., seem to count far more in ball throwing than in performances in educational tests. Distance throw and the other physical stunt tests were given, therefore, not only to aid in diagnosing the physical condition of the children but to furnish relatively simple situations in which to attempt to trace the effects of various factors. They differ among themselves in complexity. Hence the results of the series of measurements present a narrow but consistent field in which to attempt to trace the operation of persistent factors.

Physical development and condition are thought by some to influence intellectual achievement, especially scores in tests. One problem, therefore, was to trace the effect of heredity, maturity, and training in the physical field upon the results secured in achievement tests, particularly rate tests. Three types of tests, designed to mea-

sure, among other things, rate of reading, and rate of adding, and rate in simple copying exercises, were given to furnish further material for such analysis.

For formal measurement of intelligence, two standard tests were used: the National Intelligence Test and the Army Test, as modified for use in Detroit. The first furnish a standard group measure of intelligence and the second a basis for comparison with children in other city schools.

The results of training were measured with the Stanford Achievement Tests, a battery of carefully prepared and standardized performance scales in reading, arithmetic, spelling, and other school subjects. The scores make comparisons with other cities possible, as well as furnish a composite measure of educational achievement, or educational age. Here, too, the series of educational results was reduced to a single index figure.

A music appreciation test was given at the beginning and end of the semester to investigate the effect of a highly specialized capacity, namely, appreciation of music.

HANDWRITING. The field is one in which many persons find it easy to change their method of work. Three specimens of handwriting were secured under different types of conditions. These scores were valuable not only as material for a study of relation of rate to quality, but as a field in which to trace the effect of the physical elements.

ADDING. A special adding test was used to parallel and confirm or refute the conclusions from the handwriting test, both in regard to change of method and to the effect of the motor element. Four trials of the test were given under varying conditions.

MOTOR-MENTAL. A series of eight tests designed to measure rate of motor activity and to reveal the effect of the addition of an extraneous thought element to the motor complex. They were thought to bear a close relation to both the physical and the intelligence factors.

ORAL READING TEST—GRAY. The test was given primarily because it was believed the scores secured are related to maturity and physical ability, as well as to reading ability. The test was expected to serve, therefore, as an illustration of the way in which scores may really represent abilities quite different from those they were intended to measure.

SILENT READING TESTS—DETROIT RATE TEST. The test was chosen, not so much to measure reading ability as to furnish an opportunity to trace the influence of the physical factors in other fields.

COURTIS READING TEST, KITTEN SERIES. This test furnishes a similar illustration, but under quite different circumstances. It is believed that in the lower grades the scores reflect reading ability of a simple type, but that in the upper grades the test measures mainly motor ability. In this case, a check motor test was given to furnish additional data on which to base final conclusions.

THORNDIKE-McCALL READING TEST. The motive for selection in this case was quite different. There has been some evidence offered tending to show that this test is more nearly a measure of intelligence than of reading ability. With the data from the other tests available, it was expected that an analysis would reveal what factors really do determine the score.

SPELLING TESTS. Special tests were constructed from the Ashbaugh Spelling Scale to afford material for another validation study. Two tests were given, one covering the present grade of the children and the grade below, the other the present grade and the grade above. These, together with the dictation test in the Stanford series, furnished the necessary data.

COMPOSITION TESTS. Efforts were made to secure additional data in regard to parents, homes, ambitions, and ideals, through a series of four compositions on the themes—My Ideal Mother, How I Spent My Saturday, How I Spent My Sunday, and What I Expect To Do When I Grow Up. These compositions were also scored for rate and quality and furnish other subject matter for analysis.

EMOTIONAL SURVEY. Teachers were furnished with detailed descriptions of the way children behave in a number of emotional situations, and were asked to record their observations and judgments in regard to each child in their care. More than one teacher rated most of the children. These ratings were summarized and reduced to an index figure which expressed crudely, but definitely, the degree of emotional control developed by a given child. These indices were used to trace the effect of emotional control upon success.

SOCIOLOGICAL SURVEY. For children ten, eleven, and twelve years old, an investigator was sent to the home to interview parents and to secure information bearing upon the health and life interests of the child. Incidentally, the investigator recorded his observations in regard to a number of significant details that serve to fix the status of the home on the social scale. This phase of the investigation proved most costly and difficult to pursue; hence it was limited to the largest age groups.

MEASUREMENT OF SUCCESS. Since one of the important objectives of the investigation was to determine why some children succeed and others fail, various records of success were secured. For one thing, teachers were asked to arrange the names of their children in order of their estimated natural capacities, and then to indicate which ones had measured up to capacity and which had fallen behind. For a second measure, the half-terms of progress in school for each child were divided by the half-years spent in school. The resulting decimal fraction it was thought, might furnish a rough measure of capacity. For a third measure, the actual school marks as reported by the teacher at the end of the year were taken. Records were also made of physical defects, absence, and other items taken from the school files. Finally, scores in the routine standard tests in use in the city were compared with the results of the special tests. Taken altogether, the series of results for each individual furnished a rather complete and objective record of his development and achievements.

II

Sample Tests and Records

Because a knowledge of test elements is frequently essential to correct interpretation of test results, it seems best to insert here samples of tests and records. However, to save space, only titles and the first and last portions of tests are shown. Most of the illustrations are considerably reduced in size. Those who need to study the tests in detail are referred to the tests themselves and to the manuals of instruction issued by their publishers. The illustrations given here are for purposes of general orientation only.

INDEX	PAGES
Stanford Achievement Test.....	212
National Intelligence Test.....	215
Detroit Army Intelligence Test.....	217
Thorndike-McCall Reading Scale.....	218
Detroit Reading Test.....	219
Reading Test No. 2. Kitten Series.....	220
Detroit Addition Test.....	221
Handwriting Test.....	221
Motor Mental Tests.....	222
Social Survey Report Blank.....	223
Emotional Survey, Instructions for Rating.....	225
Record Cards, Social and Emotional Surveys.....	229

Stanford Achievement Test

By TRUMAN L. KELLEY, GILES M. RUCH, and LEWIS M. TERMAN

ADVANCED EXAMINATION: FORM A

FOR GRADES 4-8

TEST 1. READING: PARAGRAPH MEANING

Sample: Dick and Tom were playing ball in the field. Dick was throwing the ball and
was trying to catch it.

Write **JUST ONE WORD** on each dotted line.

- 1 Fanny has a little red hen. Every day the hen goes to her nest and lays an egg for Fanny to eat. Then she makes a funny noise to tell Fanny to come and get the
- 2 A kitten can climb a tree, but a dog cannot. This is very lucky for Nellie's kitten. Every time Joe's big dog comes along the kitten climbs a tree and the cannot follow.
- 29 "Naïve" and "unsophisticated" are frequently confused. The former suggests a type of behavior which is artless, spontaneous, and free from the restraints of custom. The latter implies fully as great lack of knowledge of social usage, and, in addition, conduct which is primitive and perchance inelegant. Thus, the youth was the first to enter the car, and his little sister warmly kissed him in the presence of the king. We may also say that a country boy is with respect to city life and customs.

Test 1. Number right $\times 2 =$ Score

TEST 2. READING: SENTENCE MEANING

Samples: Can dogs bark? Yes No

Does a cat have six legs? Yes No

Read each question and draw a line under the right answer.

- | | | | | |
|----|--|------------|----|----|
| 1 | Is milk white? | <u>Yes</u> | No | 1 |
| 2 | Do we sleep in beds? | <u>Yes</u> | No | 2 |
| 39 | Does speaking with brevity necessarily mean that one is peevish? | <u>Yes</u> | No | 39 |
| 40 | Are chimes ever played in a cathedral? | <u>Yes</u> | No | 40 |
| 79 | Is a certain amount of bias necessarily degrading? | <u>Yes</u> | No | 79 |
| 80 | Are convictions usually made on insufficient evidence? | <u>Yes</u> | No | 80 |

Number right

Number wrong

Test 2. Score (subtract)

TEST 3. READING: WORD MEANING

Samples: Bread is something to catch drink eat throw wear
 A robin is a bird cat dog girl horse

In each sentence draw a line under the word that makes the sentence true.

- 1 March is the name of a day food month week year..... 1
 2 A fat person is always bad blue cold heavy little..... 2
 84 A reprobate is one who is very cowardly ugly wealthy wicked youthful..84
 85 Candid means illegitimate impeccable imperious incisive ingenuous.....85

Test 3. Score

TEST 4. ARITHMETIC: COMPUTATION

Get the answers to these examples as quickly as you can without making mistakes.

Look carefully at each example to see what you are to do.

Begin here.

(1)

(2)

(3)

(4)

(5)

$$3 + 2 =$$

$$3 + 4 =$$

$$\begin{array}{r} \text{Add} \\ 2 \\ 5 \\ \hline \end{array}$$

$$\begin{array}{r} \text{Add} \\ 7 \\ 4 \\ \hline \end{array}$$

$$\begin{array}{r} \text{Add} \\ 13 \\ 2 \\ \hline \end{array}$$

(45)

(46)

(47)

$$67.36 + \frac{2}{3} =$$

Multiply

$$\begin{array}{r} 4 \text{ gals. } 3 \text{ qts. } 1 \text{ pt.} \\ 4 \\ \hline \end{array}$$

Express as a decimal
to three places

$$\frac{29}{64} =$$

Test 4. Number right $\times 4$ = Score

TEST 5. ARITHMETIC: REASONING

Find all the answers as quickly as you can.

Write the answers on the dotted lines.

Use the blank sheets of paper to figure on.

Begin here.

- 1 How many are 3 eggs and 2 eggs? Answer.....
 2 Mary is 7 years old. How old will she be in 3 years? Answer.....
 39 It costs 43 cents to send a 10-pound parcel post package from New Orleans to Dallas. What will it cost to send an 8-pound package if the cost is 3 cents more on the first pound than on additional pounds? Answer.....
 40 If the hour hand of a clock is 3 inches long and the minute hand is 4 inches long, how far apart are the tips of the two hands at 9 A.M.? Answer.....

Test 5. Number right $\times 4$ = Score

TEST 6. NATURE STUDY AND SCIENCE

Samples: The number of cents in a dollar is 200 100 300Our rain comes from the clouds moon stars

Draw a line under the word that makes the sentence true.

Begin here.

- 1 Thanksgiving comes in July January November..... 1
 2 The earth is shaped most like a baseball football pear..... 2
- 94 The most expensive of these rugs is Axminster Brussels Oriental..... 94
 95 Fondant is a kind of candy meat salad..... 95
- Number right
- Number wrong + 2 =
- Test 6. Score (subtract)

TEST 7. HISTORY AND LITERATURE

Draw a line under the word that makes the sentence true.

- 1 An elf is a kind of animal brownie dragon..... 1
 2 "The Glass Slipper" reminds us of Ali Baba Cinderella Goldilocks..... 2
- 94 The Mohammedan Bible is the Bagavad-gita Koran Zend-Avesta..... 94
 95 The singular of "are" is is was were..... 95
- Number right
- Number wrong + 2 =
- Test 7. Score (subtract)

TEST 8. LANGUAGE USAGE

Samples

Apples is are good.He told telled me.

- 1 I calculate expect to go soon.
 2 Last year uncle ^{gave} give me a pair of skates.
 3 His leg was ^{broke.} broken.
 4 They have ^{gone} went to town.
 5 He isn't ^{any} no better than you.
- 59 He has no fear; nothing can ^{confuse} daunt him.
 60 Is that ^{he?} him?

Number right

Number wrong

Test 8. Score (subtract)

TEST 9. DICTATION EXERCISE

Test 9. Full score for easier sentences not dictated.....

Number right in sentences dictated.....

Sum..... $\times 2 =$ Score

NATIONAL INTELLIGENCE TESTS

SCALE B — FORM 2

Prepared under the auspices of the National Research Council by M. E. Haggerty,
L. M. Terman, E. L. Thorndike, G. M. Whipple, and R. M. Yerkes, *Chairman*

Test 1

Do this work in arithmetic as quickly as you can without making mistakes. Try each example as you come to it. Look carefully at each one to see what you are to do.

Begin here	(1) Add	(2) Multiply	(3) Subtract	(4) Divide	(5) Add	(6) Multiply
	$\begin{array}{r} 2 \\ 3 \\ \hline \end{array}$	$2 \times 4 =$	$\begin{array}{r} 4 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 3 \overline{)6} \end{array}$	$\begin{array}{r} 17 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 27 \\ 3 \\ \hline \end{array}$

(20)

 $16\frac{2}{3}\%$ of 120 =

(21)

Multiply

62 lb. 7 oz.

 $\begin{array}{r} 5 \\ \hline \end{array}$

(22)

Subtract

 $9.4 - 4.00083 =$

Test 2

In each sentence draw a line under the one word that makes the sentence true, as shown in the samples.

SAMPLES { Sheep eat mostly nuts grass fruits bread
The number of cents in a dime is 2 5 10 25

Begin here

- 1 The apple grows on a bush shrub tree vine 1
2 Wine is generally made from apples grapes oranges peaches 2
- 39 "Little Lord Fauntleroy" was written by Barrie Burnett Burns Scott 39
40 The days and nights are nearest equal in January June March May 40

Test 3

Draw a line under the right answer to each question. Do as many as you can.

SAMPLES { Can cows eat? Yes No
Do stones swim? Yes No

Begin here 1 Do children play? Yes No
2 Has a horse five legs? Yes No

- 39 Do illiterate men read textbooks? Yes No
40 Is theosophy a religious cult? Yes No

Test 4

Read carefully the first three words in each line. Then read the last four and draw a line under the right one.

SAMPLES { shoe — foot ——— hat — coat nose see head
sky — blue ——— grass — grows summer green tall
bird — sing ——— dog — tail bark walk kennel
bird — fly ——— dog — tail bark walk kennel
dress — cloth ——— hat — head wear band straw

Begin here

- 1 gun — shoot ——— knife — run cut hat bird 1
2 knit — yarn ——— sew — thread needle woman spool 2
- 31 like — kindness ——— hate — love men insults angry 31
32 doubtful — probable ——— probable — certain perhaps near far 32

Test 5

If the two things in a pair are the same, write S. If they are different, write D. Do each one as you come to it.

Begin here 430 430

052 055

5628913653 5628813653

6558348219 6558348219

DETROIT PUBLIC SCHOOLS

Department of Special Education

Psychological Clinic



10. A B C D E F G H I J K L M N O P

TEST 2

Get the answers to these examples as quickly as you can.
Use the margin to figure on if you need to

- SAMPLES { 1 How many are 5 men and 10 men?.....Answer (15)
2 If you walk 4 miles an hour for 3 hours, how far do you walk?.....Answer (12)

- 1 How many are 30 men and 7 men?.....Answer ()
2 If you save \$7 a month for 4 months, how much will you save?.....Answer ()
20 A commission house which had already supplied 1897 barrels of apples to a cantonment, delivered the rest of its stock to 29 merchants. Each merchant received 54 barrels. What was the total number of barrels supplied?.....Answer ()

TEST 3

This is a test of common sense. Below are ten questions. Three answers are given to each question. You are to look at the answers carefully; then make a cross in the square before the best answer to each question, as in the sample:

- | | |
|--|--|
| <p>1 If it rains when you are starting to go for the doctor, what should you do?</p> <p>☐ stay at home</p> <p>☐ take an umbrella</p> <p>☐ wait until it stops raining</p> | <p>10 A country should have many railroads, because</p> <p>☐ they make it easy to travel and carry goods</p> <p>☐ they are good for the steel business</p> <p>☐ they decrease the price of food materials</p> |
|--|--|

TEST 4.

Notice the sample sentence:

People hear with the EYES EARS NOSE MOUTH

The correct word is EARS because it makes the truest sentence.

In each of the sentences below, you have four choices for the last word. Only one of them is correct. In each sentence draw a line under the one of these four words which makes the truest sentence. The two samples are already marked as they should be.

SAMPLES { People hear with the eyes ears nose mouth
France is in Europe Asia Africa Australia

1. The meat of cows is called veal venison beef mutton
2. Plums grow on a shrub vine bush tree
39. Headquarters of the Kodak Co. are in Buffalo Camden Rochester Battle Creek
40. Bombay is in Japan China Egypt India

DETROIT PUBLIC SCHOOLS

TEACHERS COLLEGE

Rate of Reading Test, Form B

Instructions: Read each paragraph, then draw a line under the one answer of the five which is the right answer to the question.

ILLUSTRATION

When do the little busy bees buzz around all day, flying from flower to flower gathering honey to carry back to their hives?

At night, one Christmas day, in winter, on
January first, in summer.

"In summer" is underlined because it is the right answer.

When the signal to start is given, not before, begin to answer the questions below. Work as rapidly as you can *but do not hurry*. Only the answers right will count. Answer as many questions as you can in the time allowed. There are four pages of test questions.

1. John and Mary had a pet chicken of which they were very fond. One day neighbor Brown gave them another. Which word tells how the children felt?

Sad, sorry, cheerless, afraid, happy

28. Walter's hair had grown very long. His mother told him he must have it cut next day. To whom did he go to have his hair cut?

Doctor, lawyer, baker, butcher, barber

29. Most boys like to play the game in which small round glass balls are used. What do you think is the name of this game?

Football, hide and seek, kites, marbles, cards



DETROIT PUBLIC SCHOOLS
Department of Educational Research
Silent Reading Test No. 2

READING
SERIES R

Form 2

The Kitten Who Went to a Picnic

Directions

This test is given to see how well you can read to yourself. When the signal to start is given, open the cover and begin to read the story. Read *silently*, and only as fast as you can get the meaning; for when you have finished, you will be asked to answer questions about what you have read. You will be marked for both how much you read and how well you understand it, but it is better to get the meaning of the story than to read too fast.

When the examiner says "Mark," draw a line around the last word read, and keep right on reading. If you should finish before the examiner says "Stop," close your paper and wait quietly until the others finish.

The Kitten Who Went to a Picnic.

7 John, the hired man, was washing the automobile
15 down by the barn. He was getting it ready for
23 a picnic. Bobby and Kitten were watching him.
33 They liked to see the water clean the mud off
43 the wheels.

Directions (continued)

When the examiner is sure you understand what to do, he will give you the signal to start. Open your papers and begin to work. Answer as many questions as you can in the time allowed, but it is better to have the answers right than to try a great many questions. *Do not guess at the answers.* If you do not remember what the story said, read the *paragraph just above the questions* again and again until you find the right answer. There are four pages of questions. As soon as you finish one, begin another. When the examiner says "Mark," draw a circle around the number of the question on which you are working and keep right on working. Stop at once when the examiner says "Stop."

Up at the house everyone was busy, too. This year the whole Bennett family was going to spend the Fourth of July in the woods. They planned on taking along many baskets of good things to eat.

1. Was the picnic to be held in the woods?..... Yes No
2. Were they going on the third of July?..... Yes No
3. Was the whole family going to the picnic?..... Yes No
4. Were they to take along good things to eat? Yes No
5. Was everyone out at the barn with John?..... Yes No

DETROIT PUBLIC SCHOOLS

TEACHERS COLLEGE

Test 1. Addition — Standard

Instructions. When the examiner says "Start," work as many of the examples below as you can in the time allowed, but be sure you get the right answers. Do not hurry but work as rapidly and steadily as you can work correctly.

Each time the examiner says "Mark," draw a line around the example on which you are then working, go to the next. Do not finish an example after drawing a circle around it.

17	59	71	28	37	36	14	23
86	84	93	39	45	41	76	31
97	13	17	76	12	46	22	68
88	46	71	11	64	95	99	96
25	85	33	57	78	56	56	47
<u>65</u>	<u>22</u>	<u>64</u>	<u>88</u>	<u>91</u>	<u>81</u>	<u>31</u>	<u>53</u>

Test 2. Addition — Fast

Instructions. In this test you are to add just as rapidly as you can. It is the number of examples you try that counts.

Test 3. Addition — Slow

Instructions. In this test be sure every answer is right. Work slowly and carefully, checking your work by re-adding. It is the number of answers you have right that counts.

Test 4. Addition — Standard

Instructions. In this test you are to work as you usually do. You will be marked for both speed and accuracy.

Eorm A

HANDWRITING TEST II

Instructions: When the examiner says "Start" copy as much of the following paragraph on Handwriting as you can accurately and well in the time allowed. Do not hurry. Write as you would write a letter. You will be marked for how much and how well you write.

Every time the examiner says "Mark," put an X under the last letter you have written, and keep right on writing until he says "Stop."

Good writing is that which is easily read and easily written. Slow writing has but little value, and writing which cannot be read is worthless. It means waste of time and effort on the part of the reader. On the other hand good writing saves time, has a commercial value, and is always in demand.

Form A
Motor Activity

DETROIT PUBLIC SCHOOLS
TEACHERS COLLEGE
Rate of Action Test

Form B
Mental Activity

PART I

When the examiner says start, do what he tells you to do. Work as rapidly and as steadily as you can. Your score is the number of marks made per minute.

.
.

PART II

Copying Figure Test

Copy on this paper, in the space between the lines as many of the printed figures as possible in the time allowed. Write as rapidly as possible, forming the figures no more carefully than in working examples.

2 4 9 6 7 4 2 9 7 6 6 2 9 4 7 7 2 9 6 4 2 4 9 7 6

PART III

Copying Letter Test

Copy as many letters as possible in the time allowed. Write at the highest speed at which you can make well-formed legible writing.

etaor hesin ldetc agoer

PART IV

Printing Test

When the examiner says start, print as many of the capitals below as you can in the time allowed. Print them in the same position as the copy and work as rapidly as you can.

1. R B O U S I C E M Z / O M E C 1 \ S O B A 1

DETROIT PUBLIC SCHOOLS
TEACHERS COLLEGE
SOCIAL SURVEY

Name _____

Address _____

1. Health

a. Satisfactory

b. Sickness

c. Habits

d. Comparative

e. Accidents

f. Sicknesses

g. Operations

h. Health as
a baby

i. Teeth

j. School help

2. Personal**Traits***a.* Abilities*b.* Lessons*c.* Work*d.* Ambition*e.* Vocation*f.* Helpfulness*g.* Responsibility*h.* Recreation*i.* Attitude*j.* Study*k.* Habits*l.* Social Habits*m.* Influence*n.* Sunday
School*o.* Clubs*p.* Brothers and
Sisters*q.* Family**3. Parental
Resemblance***a.* Looks*b.* Physical
Health*c.* Tastes*d.* Disposition*e.* Traits**4. Parents'
Attitude***a.* Progress*b.* Contribution*c.* Corrections*d.* School Help**5. Miscellaneous**

See page 223 for form in which the three pages here represented appeared in the record blank actually used.

Energy

A. In comparison with the other members of his class does the individual usually

1. Make only momentary, spasmodic efforts? Does he exert force weakly and soon show signs of fatigue and exhaustion?
2. Make feeble efforts for a short period before giving up?
3. Show average effort, strength, and endurance?
4. Tend to persist longer than the average? Does he give up reluctantly?
5. Have great resources of power and endurance? Is he able, and on occasion does he, exert great effort over relatively long periods? Does he release energy easily? Is he outstandingly persistent?

Emotional Response

B. In comparison with the other members of the class, is the individual

1. Unemotional; phlegmatic even in extreme situations. insensitive to stimuli which stir emotions in others?
2. Abnormally emotional, so that he gives evidence of great emotional reactions to situations which appear commonplace to others?
3. Less emotional than the average, responding only to situations and stimuli which are out of the ordinary?
4. Rather more easily stirred than the average? Are all his activities plainly accompanied by emotional reactions?
5. So well-balanced and controlled that while he is sensitive to the emotional elements in every situation, he shows only those emotions and feelings which are appropriate to the various situations which arise?

True

Detroit Public Schools

Teachers College

SURVEY OF EMOTIONAL TENDENCIES AND BEHAVIOR

Instructions. Study the following paragraphs until you are perfectly familiar with them. Study the child, observing how he behaves in various situations. Then make appropriate check marks on the record card to indicate your judgments.

Note that the judgments called for range from 1 (worst) to 5 (best) that the difficulty in making judgments comes from the fact that observed behavior is a resultant of emotional tendencies modified by training, and controlled by reason. On the record card indicate any additional information which will throw light on the child's nature or tendencies. Do not hesitate to add new items and to note any marked ideals, whether good or bad. Try to answer the questions, "How does the child act" and "Why" Leave blank all the questions for which you have no information.

One

Imagination

C. In comparison with the other members of his class is this individual

1. Unimaginative, unable to foresee events, unable to anticipate pleasures, difficulties, and dangers, and therefore unable to plan intelligently? Is he "practical", unfantastic, unromantic?
2. Deficient in imagination so that his plans, and anticipations cover only the grossest details of the immediate future?
3. Average in imagination: able to foresee, anticipate, invent as well as the ordinary person?
4. Good in planning and anticipation? Is he resourceful? Has he initiative and originality? Can he enter into stories, read or told, and make them real?
5. Exceptional in resourcefulness, originality, initiative, planning, and anticipation of the future? Is he continually inventing, designing, suggesting, contributing to group activity? Is he fanciful, romantic, philosophic?

Generalization

D. Power to benefit by experience. Does the individual

1. Make no personal generalizations? Benefit by experience very slowly through habit formation only?
2. Make few generalizations and those only of the most general kind? (Be good. Study hard.) In the main, does he adopt slogans of others?
3. Make generalizations in terms of narrow, specific situations and habits? (I concentrate. I don't waste time.)
4. Make generalizations in terms of individual method, and development only? (I have almost learned to control my temper, I never make the same mistake twice.)
5. Make generalizations in terms of general methods and of group responsibility? (Each of us ought to do his share.)

Three

E. What does he do in situations which threaten **personal danger** to self or others? Is he

1. Unmoved and undisturbed? Does he accept situation solidly and uncomprehendingly?
2. Slightly disturbed? More so as the situation is gone over again and again in imagination?
3. Greatly excited and upset? Does he make an emotional response only, become hysterical, panic, etc.
4. Excited but reasonable? Does he make some attempt to do something to avoid or avert the danger?
5. Poised, well-controlled? Does he respond by intelligent, appropriate action?

F. How does he behave in situations involving **personal discomfort**? Does he

1. Stoically endure discomfort, making no response?
2. Become slightly irritated or distressed?
3. Give expression vigorously to emotional reactions only, crying, fretting, complaining?
4. Try to keep emotions under control and to do something to remedy conditions?
5. Control his emotions and take intelligent action?

G. How does he behave when he meets a **difficulty** either in planning or in the execution of plans? Does he

1. Give up as soon as difficulty is recognized and without emotional reaction?
2. Give up easily but with expressions of distress, worry, etc.?
3. Give up reluctantly and with evidence of marked emotional disturbance; worrying, fussing, crying?
4. Tend to persist and to make an effort to control his emotions?
5. Stick to the task, controlling his emotions and making intelligent efforts to solve difficulty?

Four

H How does he behave in situations involving opportunities for novel or exciting experiences? Does he

1. Give evidence of shrinking, distaste, fear, etc., doing all he can to avoid them?
2. Exhibit considerable reluctance to accept opportunities, raising objections? Is he conservative?
3. Passively accept opportunities without emotional response?
4. Give evidence of anticipation of pleasure and enjoyment, but do little to achieve same?
5. Accept eagerly, with marked enjoyment and intelligent efforts?

I. How does he behave in situations involving recognition by others of success or failure of self? Does he

1. Exhibit perfect "passion" for being "in the lime light", whether from success or failure? Will he do anything to attract attention? Is he jealous of recognition of others?
2. Desire recognition but do little to secure it?
3. Ignore recognition of self or others because he is apparently unaffected by it?
4. Appreciate meaning of recognition to such an extent that his emotions interfere with intelligent response?
5. Appreciate recognition without being upset by it, so that he gets the greatest benefit from a situation, whether it involves success or failure?

Five

J. How does he behave in situations involving relations with others? Does he

1. Repel advances from others and seek solitude or at most the company of one or two select friends? Is he non-social?
2. Tend to avoid friendly and social relationships but yield if sufficiently persuaded?
3. Neither avoid advances from others nor make advances to them?
4. Try to establish friendly relations with others? Does he make the first advances?
5. Become unhappy unless in the company of others? Does he try to make friends with everyone? Is he markedly social in nature?

K. How does he behave in situations involving property possessions, etc.? Has he

1. A passion for possession, a great desire to get and hold everything seen to involve ownership? Does he try to exclude others from sharing in the same? Is he envious of possession by others?
2. A strong desire for acquisition and ownership but some recognition of the rights of others to own also?
3. Little sense of possession? Is he careless with property of self and others? If he is non-social, is he indifferent to money? If he is social, is he lavish in gifts to others?
4. Some appreciation of the value and use of property and of the rights of ownership, with conflicting desires for possession which interfere with completely intelligent action?
5. Intelligent appreciation of property right and needs of self and others with emotions subordinated to reason so that his actions are appropriate for the situation?

Six

Records for Social and Emotional Survey

Name _____		Block Number _____		Block Standard _____		GRADE
LAST NAME		1	2	3	4	5
		Worst				Best
P. C.	A. Energy					
	B. Emotions					
	C. Imagination					
	D. Generalization					
I. Q.	E. Danger					
	F. Discomfort					
E. Q.	G. Difficulty					
	H. Novelty					
S. Q.	I. Recognition					
	J. Social Relations					
E. Q.	K. Property					
	L. Work					
Index of Social Conditions Sociological Survey, DETROIT TEACHERS COLLEGE						NUMBER

Name _____		Block Number _____		Block Standard _____		GRADE
LAST NAME		1	2	3	4	5
		Worst				Best
P. C.	I. Type					
	II. Condition					
	III. Standard					
	IV. Furnishing					
I. Q.	V. Appearance					
	VI. Language					
E. Q.	VII. Attitude					
	VIII. Interest					
S. Q.	IX. Condition					
	X. Unity					
Index of Emotional Condition Sociological Survey, DETROIT TEACHERS COLLEGE						NUMBER

M. RELAXATION PREFERRED _____

N. MOST EFFECTIVE INCENTIVE _____

O. DOMINANT STATE _____

P. CHARACTERISTIC ATTITUDE _____

Q. TALENTS _____

R. VIRTUES OR DEFECTS _____

III

DETAILED INSTRUCTIONS FOR GIVING AND SCORING CERTAIN TESTS FOR WHICH NO STANDARD DIRECTIONS WERE AVAILABLE

NO. I—RANK ORDER OF PUPILS

DIRECTIONS TO TEACHERS.

In order to check and interpret the report on "Pupil's progress" in the light of teachers' judgment, it is requested that each teacher rank the pupils in her room in accordance with the following directions:

Under the column labelled name, arrange the names of the pupils alphabetically. In the second column, place after each name the number which indicates the rank order of that pupil in terms of his natural capacity, putting in first place the child who has the greatest degree of natural gifts, and last the pupil having the least, with the others ranged between.

In the next column, place the number which indicates the rank order in terms of school achievement which you feel the child has made. The child having rank one in the second column will not necessarily be first in the third column. When these columns are completed, place in column four the explanation of the discrepancy in ranks, if any.

PROGRESS AND SUCCESS SURVEY

Name of Teacher Grade

School Room

	Name of Pupil	Rank in Order of Natural Capacity	Rank in Order of Achievement Progress	Explanation of Difference
1				
2				
3				
4				
39				
40				

NO. II—DIRECTIONS FOR WRITING TEST

TIME ALLOWANCES.

3B-5A—Three one-minute intervals.

6B-8A—Three one-half-minute intervals.

DIRECTIONS.

Supply each child with three sheets of paper and one copy of the test. Read your directions. Be careful about timing. Use only a watch with a second hand. When all are ready, say

"We are going to have three tests in writing today. Put your name, grade, and school on your paper just as you usually do. Mark this sheet Test I. Try to write as rapidly as you can write well." (These are the testing conditions designated as "normal" or standard.)

"Get ready," "Start," (60 seconds), "Mark," (60 seconds), "Mark," (60 seconds), "Stop." "Find your rate and record it on your paper."

"Now prepare your second paper for a test. Mark this Test II."

"This is a speed test, a race. Never mind your quality; see how many letters you can make per minute. Some children are not able to change their habits. When they are told to go fast, they write just as they did in the first test. How many understand that they are to make just as many letters as possible in the time allowed? Of course, I must be able to read what you write, but I expect your quality will be poorer, because you are going to see how fast you can move your fingers." (These are the testing conditions designated as "fast.")

"Get ready," "Start," (60 seconds), "Mark," (60 seconds), "Mark," (60 seconds), "Stop." "Record your rate."

"Prepare your papers for the third test. This is another race, but one of just the opposite kind. You are to give me your very best writing. Go slowly and carefully. Make every letter the right shape. Form each one with care. Keep the letters on the line. Leave plenty of space between words and letters. Watch your margins. In other words, give me a sample of your best slow careful writing." (These are the testing conditions designated as "slow.")

"Get ready," "Start," (60 seconds), "Mark," (60 seconds), "Mark," (60 seconds), "Stop." "Record your rate."

"Be sure your name and grade is on each paper."

DIRECTIONS FOR SCORING HANDWRITING

Papers were sorted by grades. There were three papers for each child.

Scoring was done by two scorers.

Scorer A would distribute all papers from one grade into four groups by chance, in order to scatter the papers for any one child. The four groups were then recombined into one group.

Scorer A would arrange the papers into about five rough groups, according to quality.

Scorer A would then arrange all the papers in rank order, from good to bad, in one long row on tables.

Scorer B would check the rank order, removing any papers he felt to be out of place.

Both scorers would then consider the papers so removed, replacing them to the satisfaction of both.

Then both together would assign by means of Ayres Scale, quality values to all papers, writing the value upon the back.

Scorer A would then group papers alphabetically, placing the three papers for each individual together.

Scorer B would then score the papers as follows:

The three papers for any individual were spread out and arranged in order of merit of handwriting.

Quality values were assigned to each by reference to the scale, and placed on the face of the paper.

These values were then compared with the values on the back to see two things:

1. Whether the order of merit was in agreement with the order as determined from the scores on the back. If not, the papers were thrown out to be scored by the other scorer.

2. Whether the values agreed with the values on the back.

If they varied more than 10 scale points, they were thrown out to be scored by the other scorer.

In case both scorers disagreed with the value on the back, a third scorer judged the papers.

NO. III.—DIRECTIONS FOR KITTEN READING TESTS

TEST I. RATE OF READING.

1. Distribute tests and have blanks filled out.

2. Read directions aloud, adding, "Whenever I say 'mark,' put a cross under the word you are reading, but go right on until you are told to stop."

In grades 3B through 5A, use three one-minute intervals.

6B through 8A, use three one-half-minute intervals.

Say "Mark" at the end of first and second intervals, and "Stop" at the end of the third.

TEST 2. SILENT READING, PART I.

1. Pass out tests, being sure child is provided with a pencil.

2. Have blanks at top filled out.

3. Read directions aloud to the children.

4. Say "Start." Allow exactly three minutes, saying "Mark" at the end of each 30 seconds.

Have pupils close books, holding them with the front page toward them. Say "Turn your paper like this" (turn end for end). Read instructions aloud. Have children try to answer questions 71-75. Continue directions.

When all are ready, say "Start." Allow exactly five minutes for all grades. Say "Mark" at the end of each minute, and "Stop" at the end of the fifth minute.

TEST 3.

Use Silent Reading, Check Sheet.

Read the directions aloud, adding, "Be sure you look at the word at the beginning of the line each time, so that you will make no mistake."

Allow $1\frac{1}{2}$ minutes, saying "Mark" at the end of each 30 seconds.

Have name and grade written on each sheet and collect with this sheet inside the other.

NO. IV.—WORDS FOR TEACHER SPELLING TEST

Taken from the Detroit and Ashbaugh Scales, to test each child twice on the words for his grade, once on the words of the grade below, and once on the words of the grade above. Words are arranged in order of difficulty.

Give the tests as follows

Grade.....	3	4	5	6	7	8
Test.....	1	1	2	3	4	5
		2	3	4	5	6

1. Have words numbered.

2. Allow a few minutes break between tests.

3. Have pupils put name, grade, and school on their papers.

LIST OF WORDS

Test 1 Grade 3-4	Test 2 Grade 4-5	Test 3 Grade 5-6
skate	east	depot
carpet	wish	quality
again	date	braid
pair	bear	manufacturer
chain	start	manager
early	own	author
brown	picture	shipped
catch	corner	awful
writing	few	stomach
calf	all right	bracelet
picnic	built	bargain
stories	ladies	ninety
cousin	collar	knowledge
free	violin	case
bill	studied	appear
wood	telephone	busy
yet	nephew	agreeable
leave	furniture	connect
brought	straight	neighborhood
dollar	safety	service
copy	manual	medicine
finish	favorite	grammar
caught	fierce	envelope
piano	ninety	glorious
language		excellent
Test 4 Grade 6-7	Test 5 Grade 7-8	Test 6 Grade 8 alone
check	appreciate	officer
however	losing	system
proper	crochet	national
loyal	disappoint	presented
contain	probably	assist
captain	advertisement	avoid
factories	campaign	article
avenue	attendance	evidence
special	immense	concern
vacant	recommend	suggest
nervous	mortgage	publication
possibly	physician	combination
cushion	occasionally	February
possession	minute	forenoon
original	flavor	certificate
absolutely	prompt	organization
disease	ability	especially
catalogue	university	preliminary
thorough	proposition	exhibition
disappointed	weigh	sincerely
privilege	genuine	exhausted
representative	decision	receipt
analysis	lieutenant	recommendation
efficient	referred	pneumonia
accommodate		disappointment

NO. V.—DIRECTIONS FOR MOTOR MENTAL TESTS

The tests we are going to try today are to show us how fast we can do the things we are asked to do.

Notice Test No. 1. Here are rows of dots. Our task is to see how rapidly we can strike these out. Be sure to start at the left and go right across the line, like this (show on board). Try on the sample. Try to strike the dot, but do not take so much pains that you go slowly. Whenever I say "mark," put a cross just where you are working.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 2. Here we are going to see how fast we can copy figures. When I say "mark," put a cross under the figure you are writing. Try the sample.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 3. Here are some nonsense words. This time, we are to see how rapidly we can copy letters. You are to mark the same as before. Try the sample.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 4. Here are letters printed out of their right position. You are to copy these as nearly as you can. Do not turn your papers. Try the sample.

3 two-minute intervals in grades 3 to 5.

3 one-minute intervals in grades 6 to 8.

TEST 2

This is the same blank you had before, but we are going to do different things. We will see what happens when we must think about what we are doing. Look at Test 1.

This time, we will make a pattern of each five dots. Start at the bottom, like this (show on board).

Keep the dots near the middle of your line. Try the sample.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 2. This time, we are going to write figures, but instead of copying what is here, we will look at each figure and think what number follows it and write that figure. If you find a 9, you will write a 0. Try the sample.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 3. This time, we will look at the letters, think what letter precedes it in the alphabet, and write that letter. If you find a, you will write z. Try the sample.

3 one-minute intervals in grades 3 to 5.

3 one-half-minute intervals in grades 6 to 8.

Look at Test 4. This time, you are to look at the letters and then print them just as nearly the opposite from what it now is. If it is right side up, print it so it will be wrong side up. If it is wrong side to, print it turned about. Whenever you find an O, an I, or an X, a letter which looks the same either way, look down to the line below and print the letter you find there, whatever it is. Try the sample.

3 two-minute intervals in grades 3 to 5.

3 one-minute intervals in grades 6 to 8.

NO. VII.—DIRECTIONS FOR DETROIT ARMY TEST (Forms B, F, G, H)

I am going to place (or have placed) one of these papers on your desk. It will be placed here, this way. Please do not open it or write on it until you are told to.

(Papers distributed.)

Write your name after the word "Name," both your first and last name. Take your time and write it as plainly as you can. After the word "Grade," write the grade you are in, using both the number and letter like this, "A 6." After the word "Age," write the number of years old you are. Between "Years" and "Months," write the number of months it has been since you have had a birthday. (Explain by using example.) After the word "School," write the name of the school you are attending. When you have finished, put your pencil down.

This is a test to find out how well you think; how well you can remember; and how well you can do exactly what you are told to do. Some of the things will be very easy for you. Some you may find hard, but be sure to do the best you can.

First, you will be told to do something with the squares at No. 1. Afterwards, with the circles at No. 2, and so on.

You are not expected to make a perfect grade or to finish everything in the test in the time allowed. You must be sure to stop right away when I say "Stop," and do not go to work again until I say "Go," which will always be your signal to begin. Do not look around the room during the test. Of course, we want to find how well each one can do alone, so don't watch anyone else; keep your eyes on your own work. During the test, do not ask questions. Now look at your papers, just below where you have been writing.

Look at the squares at No. 1. When I say "Go," put a cross—in smallest—square. "Go." (FIVE SECONDS.)

Look at No. 2. Notice the five empty circles. When I say "Go," put a cross in the first circle and the figure 1 in the third circle. "Go." (FIVE SECONDS.)

Look at No. 3. Notice the eight empty circles. When I say "Go," draw a line from the first circle to the fourth circle, which will pass up over the second circle, but down under the third circle. "Go." (TEN SECONDS.)

Look at No. 4. Notice the square and the triangle. When I say "Go," put a cross in the space which is the but not in the and a figure 1 in the space which is in the triangle and square. "Go." (TEN SECONDS.)

Look at No. 5. Notice the drawing. When I say "Go," put a figure 1 in the space which is only in the and a figure 2 in the space which is in the and . "Go." (TEN SECONDS.)

Look at No. 6. Notice the numbers. When I say "Go," cross out every number which is more than 60 but less than 70. "Go." (FIFTEEN SECONDS.)

Look at No. 7. Notice the three circles, and the three words. When I say "Go," put in the first circle the first letter of the first word; in the second circle, the second letter of the second word; and in the third circle, the last letter of the third word. "Go." (TEN SECONDS.)

Look at No. 8. Notice the five empty circles. When I say "Go," put in the second circle, the right answer to the question, "How many months has a year?" In the third circle, do nothing; but in the last circle, put any number that is the wrong answer to the question you have just answered correctly. "Go." (TEN SECONDS.)

Look at No. 9. Notice the three circles and the two words. If an automobile can travel faster than a horse and wagon, then when I say "Go," put a cross in the third circle, if not, draw a line under the word "No." "Go." (TEN SECONDS.)

Look at No. 10. Notice some letters of the alphabet. When I say "Go," cross out the letter just before C, and also draw a line under the second letter before H. "Go." (TEN SECONDS.)

EVERYONE STOP! Now, turn over the page for Test 2. Be sure you have the right page. Read the printed directions at the top of the page.

Examiner reads printed directions aloud, and then says "Ready; Go," for Tests 2, 3, and 4. Time allowances: Test 2, 4 MINUTES; Test 3, 1¼ MINUTES; Test 4, 3 MINUTES.

VII.—DIRECTIONS FOR COMPOSITION TESTS IN SOCIAL SURVEY

I

The assignments are to be made in four lessons at the discretion of the teacher, from 4B through 8A.

If it is possible, have children use small notebooks. Follow the general line of attack in this paper, so that results may be comparable from grade to grade.

Say, "Several boys were playing in our neighbor's yard, one day last week. From the amount of noise they were making, I judged they were having a glorious time. Presently, out came Mrs. A, with a large plate of cookies which she handed to her son, saying, 'Here, Ted, pass those to your friends. All boys like nice, fresh cookies.'"

"I noticed one little fellow who had a rather neglected look. After munching his share of the cookies, he remarked, 'I wish I had a mother like yours, Ted!'"

"We didn't choose our fathers and mothers, did we, boys and girls? Let us imagine today that we haven't any mother, and we have been asked to choose one. Just what kind of a mother would you choose? I don't want you to ask any questions, because I do not want to influence you in any way. Just think of all the things that would make up your ideal mother. Justify each statement you make."

II

When all are ready to work, say "I want to time you as you write. When I say 'Start,' begin to write, and whenever I say 'Mark,' put a cross under the word you are writing on and go right on. If you finish before I say stop, sit quietly until all have finished."

When all are writing, say, "Mark" at the end of two minutes, four minutes, and six minutes. Allow children to write approximately ten minutes in all.

Say, "I want to try an experiment this week, and I'm very anxious to make it a success. I can't do it alone. I wonder if some of you are willing to help me. Let me see how many are.

"How many have ever kept a diary? What do you do when you keep one? Then what is a diary?"

"That sounds as if we'd have to have books. Wouldn't it be nice if we all had little books just alike? Well, I'll make a bargain with you. I'll provide the little books, and you provide the records that are to go in them. Do you agree?"

"I must have a very complete record, in order to make my experiment a success; so on Monday, I'll ask you to write out in detail, the things you did on Saturday.

"When we write an experience composition, what are some of our standards? (Variety of sentence construction, one-phase composition.) I wonder if we should work for these two standards in a diary. What do you think about it? Justify.

"Let's make an outline on the board. If we are going to tell everything that happened on Saturday, with what shall we begin? What will be the first item in our outline?"

"(a) What time did I get up? (b) What did I do after getting up? (c) Did I breakfast with the others, or did I get up too late for that? (d) Who prepared my breakfast? (e) Who washed the dishes? (f) What did I do between breakfast and lunch time? (g) Who prepared lunch? (h) Who put things in order after lunch (cleared table, washed dishes)? (i) What occurred between lunch and supper? (j) Who prepared supper? (k) What happened between supper and bedtime? (l) At what hour did I go to bed?"

III

The same outline was used for Sunday's record.

IV

What I Want To Be When I Grow Up

Say, "One day, when I was teaching an A first class, I noticed one little fellow idling.

"Oh, Arthur!" I exclaimed, 'You will have to work hard if you wish to become a doctor, like Daddy, when you grow up.'

"I'm not going to be a doctor," he replied. 'I'm going to be a policeman, because all he does is to walk along the street swinging his club, and that doesn't require any studying.'

"A few days later, Arthur noticed a policeman writing in a notebook, so he immediately decided that was no job for him.

"I wonder how many of us expect to follow in Father's footsteps when we grow up.

"Who can tell us about the experience of Columbus in choosing his life work?

"Who can tell us of another famous man's experience?

"Who will impersonate Columbus and give his experience?

"Let us write a letter this morning, telling what we wish to become when we grow up.

"Shall I help you with the outline?"

OUTLINE

1. In what line of work Father is engaged.
2. What I think of that line of work.
3. Why I think so.
4. Whether or not I expect to do similar work.
5. Why or why not.
6. If not, what kind of work I would like to do.
7. Why {
 - easier
 - pleasanter
 - more profitable.

NO. VIII.—SOCIAL SURVEY

Suggested questions to be asked parents in making the survey of social conditions in the homes of the children in the experimental schools.

WARNING! All participating in the work of this survey are asked to use the utmost tact and discretion. While most parents are perfectly willing to give any information desired when they know that the purpose and the attitude of the investigator are worth while, every person resents meddlesome prying into his private affairs. Be sure, therefore, that you secure the cooperation of the parent, and do not push inquiries beyond the limits that he is willing to go. The major purpose of the survey will have been achieved when you see and talk with one of the child's parents, and it is better to have some of our records incomplete than to arouse any unfavorable reactions. Each one is asked to treat those interviewed with the utmost courtesy and consideration.

QUESTIONS

1. HEALTH OF CHILD.

- a. Is the health of (John) satisfactory?
- b. Is he sick occasionally?
- c. What are his habits in regard to eating, sleeping, etc.?
- d. How does his health compare with that of his brothers and sisters?
- e. What serious accidents, sicknesses, or operations has he had?

- f. Was he a healthy baby?
- g. How old was he when his first teeth began to appear?
- h. Is there anything the school can do to help in bettering his health conditions?

2. PERSONAL TRAITS OF THE CHILD.

- a. Has (John) any marked special abilities or activities?
- b. What special lessons does he take outside of school?
- c. Has he any regular work to do about the house or outside of it?
- d. What does he wish to become?
- e. What do you wish him to follow as a vocation?
- f. Is he helpful in the home?
- g. What amount of personal responsibility does he assume?
- h. What is his favorite form of recreation, as reading, movies, baseball, etc.?
- i. What is his attitude toward school work? Does he like school?
- j. Does he study much at home?
- k. Has he any bad habits?
- l. Does he like to play better with his companions than by himself?
- m. What influence do his companions have upon him?
- n. Does he like to go to Sunday school?
- o. To what clubs does he belong?

3. PARENTAL RESEMBLANCE.

- a. Which parent does he most resemble in looks?
- b. In physical health and condition?
- c. In tastes?
- d. In disposition?
- e. In any other marked personal traits?

4. PARENTS' ATTITUDE.

- a. Are you satisfied with his progress in school?
- b. What do you consider as the most valuable thing the school is doing for him?
- c. What tendencies, if any, are you trying to correct in your child?
- d. Is there anything that the school can do to help?

SOCIOLOGICAL SURVEY—HOME CONDITIONS

Observation to be made in the survey of social conditions in the homes of the children in the experimental schools.

INSTRUCTIONS: Immediately after an interview, make check marks on the child's card to record your observations and judgment. Note (1) means worst, and (5) means best. The first three ratings may be made at any time, the others only after an interview. Be careful about making ratings in the presence of the parents. For Nos. VII, VIII, IX, and X the essential observations may be entered on the survey sheet and the rating made later. In general, however, try to complete the record on the card immediately after an interview and before the next one.

RATINGS

I—Type of House.

- 1. House facing alley or two houses on lot.
- 2. Apartment house, flats without yards.
- 3. Flats with yards, double houses without yards.
- 4. Double houses with yards, single houses without.
- 5. Single house on lot, a yard.

Enter on card H for single house, A for apartment, F for flat, D for double house.

II—Conditions of Premises.

1. Yard of house untidy, unkept, scattered heaps of refuse, no evidence of care.
2. Rubbish in piles, premises untidy, but showing some signs of care.
3. Average neatness, order and care.
4. Better than average conditions, yard raked.
5. Marked evidence of care, lawn cut, no rubbish in sight, woodwork well painted. Everything in splendid condition.

Note special conditions such as moving, building going on.

III—Standard or value of home.

1. Single house \$1500 or less, apartment \$10,000 or less.
2. Apartments over \$10,000, flats \$8000 or less.
3. Flats over \$8000, double houses \$12,500 or less.
4. Double houses over \$12,500, single houses between \$5000 and \$10,000.
5. Single house over \$10,000.

Note special conditions, such as moving, building going on, etc.

IV—Furnishings of house (judgment as to wealth and taste of family).

1. Bare necessities, bare floor and walls, cheap furniture and little of it.
2. Home scantily furnished, or fully furnished but furnishings old and worn.
3. Home comfortably furnished or overfurnished in bad taste.
4. Home well furnished and in good taste. Books, music, or pictures in evidence.
5. Home elegantly furnished. Many evidences of appreciation of the artistic and beautiful in literature, music, and pictures.

V—Appearance of parents and home (judgment as to family standards of refinement).

1. Home very dirty or in gross disorder, disagreeable odors. Parents' clothing untidy, and marked evidences of very low standards.
2. Home somewhat dingy and unkept. Clothing soiled and ragged.
3. Home in average condition. Evidence of laxity, but on the whole comfortable. Standard of cleanliness, order, and taste maintained.
4. Condition better than average, clear evidence of much care and attention.
4. Home in perfect order. Marked cleanliness. Parents appropriately and carefully dressed. Good manners.

VI—Language conditions in the home.

1. Both parents speak little or no English.
2. One or both parents speak broken English.
3. English used by parents of average quality.
4. Better than average English spoken.
5. English used of exceptional quality in vocabulary and style.

VII—Attitude of Parents toward Child.

1. Expects child to work for parents' benefit. Antagonistic to American ideals.
2. Accepts American ways reluctantly. Keep child out of school to help in home. Believe children should be made to work.
3. Accepts general responsibility for child's welfare, but feels duty is done when child is housed, clothed, fed, and sent to school.
5. Accepts full responsibility for child's welfare and development, knows intimately the child's ambitions.

VIII—Interest of Parents in School Work.

1. Opposed to education. Think child ought to be at work, resentful.
2. Believes in some education. Somewhat reluctant about sending child to school.
3. Believes in education. Content to shift responsibility to school. Punishes child if he does not do well.
4. Accepts some responsibility. Knows principal and has been to school. Knows grade child is in.
5. Accepts full responsibility and tries to cooperate with school. Visits school and knows teacher. Quite alive to aims and purposes of school efforts.

IX—Conditions of Life in Home.

1. Marked evidence of contention and unhappiness. Fear, quarreling, and deceitfulness are much in evidence.
2. Evidence of some unhappiness and contention in the home. Child and parents out of sympathy.
3. No evidence of either happiness or unhappiness.
4. Indications of happy, affectionate family life.
5. Evidence of exceptional home life. Parents and children very congenial. Pleasures and relaxations shared. Happy understanding very marked.

X—Unity in Home.

1. Children living with strangers.
2. Children living with relatives.
3. Children living with one parent. Parents divorced or re-married.
4. Children living with one parent, other dead.
5. Children living with both parents.

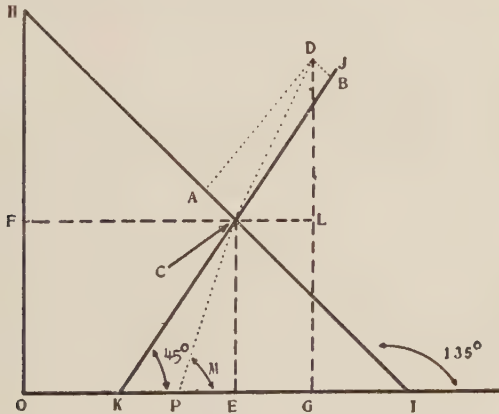
IV

Derivation of the Formula for Determining the Amount of Effort and Adjustment Involved in a Change of Performance from Point $X_1 Y_1$ to Point $X_2 Y_2$ in the Rate-Quality Field

Based on the assumption that rate and quality units have been reduced to equivalence (where necessary), so that the adjustment line is a line through the point $X_1 Y_1$ making an angle of 135° with the rate (X) axis, and the effort line is a line through the point $X_1 Y_1$ making an angle of 45° with the rate (X) axis.

In the figure let

- | | |
|---|---------------------------|
| OI be the rate axis | HI be the adjustment line |
| OH be the quality axis | JK be the effort line |
| C be the point $X_1 Y_1$ | DB be perpendicular to JK |
| D be the point $X_2 Y_2$ | DA be perpendicular to HI |
| M be the angle made by the line DC with the rate axis = angle DCL | |



Then

$$\text{Angle JCH} = 90^\circ$$

$$X_1 = CF \quad X_2 = FL \quad Y_1 = CE \quad Y_2 = DG$$

$$CL = (X_2 - X_1) \quad DL = (Y_2 - Y_1)$$

$$\text{Angle BCD} = M - 45^\circ$$

$$BC = DA = \text{the effort component of CD}$$

$$AC = DB = \text{the adjustment component of CD.}$$

Problem:

To find the value of BC and AC in terms of X_1, Y_1, X_2, Y_2 .

Solution for effort factor, BC

$$BC = CD \cdot \cos (M - 45^\circ) \dots \dots \dots (1)$$

$$= CD (\cos M \cos 45^\circ + \sin M \sin 45^\circ) \dots \dots \dots (2)$$

$$\text{But } \cos 45^\circ = \sin 45^\circ = \sqrt{.5}$$

Removing the common factor from (2),

$$BC = \sqrt{.5} CD (\cos M + \sin M) \dots \dots \dots (3)$$

$$(\cos M + \sin M) = \frac{1}{\tan M \sqrt{1 + \cot^2 M}} + \frac{\tan M}{\sqrt{1 + \tan^2 M}}$$

$$= \frac{\sqrt{1 + \tan^2 M} + \tan^2 M \sqrt{\frac{1 + \tan^2 M}{\tan^2 M}}}{\tan M \sqrt{\frac{1 + 2 \tan^2 M + \tan^4 M}{\tan^2 M}}}$$

$$= \frac{1 + \tan M}{\sqrt{1 + \tan^2 M}} \dots \dots \dots (4)$$

$$\text{But } \tan M = DL/CL \dots \dots \dots (5)$$

$$\therefore \frac{1 + \tan M}{\sqrt{1 + \tan^2 M}} = \frac{1 + \frac{DL}{CL}}{\sqrt{1 + \frac{DL^2}{CL^2}}} = \frac{CL + DL}{\sqrt{CL^2 + DL^2}} \dots \dots (6)$$

$$\text{Also } CD = \sqrt{CL^2 + DL^2} \dots \dots \dots (7)$$

Substituting (6) and (7) in (3)

$$BC = \sqrt{.5} \sqrt{CL^2 + DL^2} \cdot \frac{CL + DL}{\sqrt{CL^2 + DL^2}}$$

$$= \sqrt{.5} (CL + DL) \text{ or } \dots \dots \dots (8)$$

$$= \sqrt{.5} [(X_2 - X_1) + (Y_2 - Y_1)] \dots \dots \dots (9)$$

the desired value of the effort factor.

Solution for adjustment factor, CA

$$\begin{aligned}
 AC = DB &= \sqrt{\overline{CD}^2 - \overline{BC}^2} \text{ or substituting 7, and 8.} \\
 &= \sqrt{\overline{CL}^2 + \overline{DL}^2 - .5 (CL + DL)^2} \\
 &= \sqrt{\overline{CL}^2 + \overline{DL}^2 - .5 \overline{CL}^2 - CL \cdot DL - .5 \overline{DL}^2} \\
 &= \sqrt{.5 \overline{CL}^2 - CL \cdot DL + .5 \overline{DL}^2} \\
 &= \sqrt{.5} (CL - DL) \text{ or} \\
 &= \sqrt{.5} [(X_2 - X_1) - (Y_2 - Y_1)] \text{ or the desired} \\
 &\quad \text{adjustment factor} \dots\dots\dots (10)
 \end{aligned}$$

V

Equivalence of Prediction Formulae and Value of Constants¹

Proof that a prediction formula derived by the piecemeal method is equivalent to the corresponding formula (regression equation) derived by conventional methods. Determination of the values of the constants in the piecemeal formula.²

Given X_1 and X_2 , independent variables, and Y , a variable dependent upon variables X_1 , X_2 , among others, the prediction formulae, assuming rectilinearity of relationship, are:

Piecemeal: $Y = M_y + C_1(X_1 - M_{x_1}) + C_2[(X_2 - M_{x_2}) - C_3(X_1 - M_{x_1})] \dots \dots \dots A$

Conventional: $Y = c + b_1 X_1 + b_2 X_2 \dots \dots \dots B^3$

Problem:

To prove that equation A is equivalent to equation B, and to determine the values of C_1 , C_2 and C_3 in conventional terms.

PROOF

Formula A may be written

$$Y - M_y = C_1(X_1 - M_{x_1}) + C_2[(X_2 - M_{x_2}) - C_3(X_1 - M_{x_1})]$$

Substituting deviation symbols for each expression of the form (Variable—Mean), the equation becomes

$$\bar{y} = C_1 \bar{x}_1 + C_2(\bar{x}_2 - C_3 \bar{x}_1)$$

Upon expanding and collecting terms this becomes

$$\bar{y} = (C_1 - C_2 C_3) \bar{x}_1 + C_2 \bar{x}_2 \dots \dots \dots C$$

Equation B may be written in deviation form also

$$\bar{y} = b_1 \bar{x}_1 + b_2 \bar{x}_2 \dots \dots \dots D$$

Equations C and D are identical in form and meaning. Therefore equations A and B, from which they are derived, are equivalent.

A similar proof may be given for the equivalence of a piecemeal formula for any number of variables.

¹I am indebted to Professor H. C. Carver for the essential elements of this proof, but he should not be held responsible for any errors which may be found in my statement of it.

²Gross scores of variables, whether actual or predicted, will be indicated by the letters representing the variables, means by m with proper subscript, deviations from means by the letters with a superior bar, standard deviations by σ with proper subscript, product moment correlation coefficients by r with proper subscripts, regression coefficients by b with proper subscripts, constants by C , c .

³Kelly, T. L. *Statistical Method*, page 283. Otis, A., *Statistical Methods in Educational Measurement*, pages, 243-4.

Value of the Constants, C_1 , C_2 , and C_3

1. Value of C_1

From comparison of equations C and D it is evident that

$$b_1 = (C_1 - C_2 C_3) \dots \dots \dots E$$

$$b_2 = C_2$$

The fundamental assumption on which the determination of constants in piecemeal derivation is based is that

$$\bar{y} = C_1 \bar{x}_1$$

Multiplying by $\bar{x}_1$ and summing

$$\sum \bar{x}_1 \bar{y} = C_1 \sum \bar{x}_1^2 \quad \text{or}$$

$$C_1 = \frac{\sum \bar{x}_1 \bar{y}}{\sum \bar{x}_1^2}$$

$$\text{Replacing } \sum \bar{x}_1^2 \text{ by } \sqrt{\sum \bar{x}_1^2} \sqrt{\sum \bar{x}_1^2}, \text{ multiplying by } \frac{\sqrt{\sum \bar{y}^2}}{\sqrt{\sum \bar{y}^2}}$$

and re-arranging terms, the equation becomes

$$C_1 = \frac{\sum \bar{x}_1 \bar{y}}{\sqrt{\sum \bar{x}_1^2} \sqrt{\sum \bar{y}^2}} \cdot \frac{\sqrt{\sum \bar{y}^2}}{\sqrt{\sum \bar{x}_1^2}}$$

$$\text{Therefore } C_1 = r_{x_1 y} \cdot \frac{\sigma_y}{\sigma_{x_1}} \dots \dots \dots F$$

2. Value of C_2

$$C_2 = b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{x_2 y} - r_{x_1 y} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \dots \dots \dots G$$

3. Value of C_3

From equation E above.

$$C_3 = \frac{C_1 - b_1}{C_2}$$

Substituting the proper values for these constants,

$$C_3 = \frac{\frac{\sigma_y}{\sigma_{x_1}} \cdot r_{x_1 y} - \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{x_1 y} - r_{x_2 y} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}}{\frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{x_2 y} - r_{x_1 y} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}}$$

or

$$C_3 = \frac{\sigma_{x_2}}{\sigma_{x_1}} \cdot r_{x_1 x_2} \dots \dots \dots H$$

ILLUSTRATIVE COMPARISON OF THE USE OF THE CONVENTIONAL AND PIECEMEAL METHODS OF DERIVATION OF PREDICTION FORMULAE IN THE ANALYSIS OF THE RELATION BETWEEN AGE AND GRADE ATTAINED IN SCHOOL

When used with identical data, the piecemeal derivation of a prediction formula yields precisely the same results, although expressed in different form, as the conventional method. The question naturally arises, "Why use an awkward and cumbersome method which at best expresses truth in relationships varying from term to term, and needing to be changed to conventional relationships to yield the information sought?"

The answer is that the piecemeal method, in spite of its limitations, has also the peculiar merit (1) of enabling the investigator to easily study the relationships between a dependent variable and each of its factors, one factor at a time, no matter how many factors may ultimately be included in the formula; (2) of pointing out just which individual data vary from the average relationship and to what degree, thus guiding the investigator in his search for causes; and finally (3), of enabling the prediction formula to be written in a form which reveals the manner and the degree to which each factor interacts with each other factor in the relationships on which the prediction is based.

Nothing can be done with the piecemeal method which may not be done with the conventional method. The piecemeal method merely provides for computations, and arrays the results in such forms that they are persistent stimuli to analytical thinking, just as the conventional procedure is a constant stimulus to thinking in mass, or average, terms. All of the arguments in favor of the use of the piecemeal method in the last analysis reduce to the one; namely, the stimulus to, and the ease of, analytical thinking. The illustration given below is to be read from this point of view.

The relationship chosen for illustrative investigation is that between the age of school children and the grade attained by them in school. The chief reason for this choice is that the dependent variable, grade attained, is an important complex resultant of the play of several general and specific factors, all of which are known, and all of which may be measured objectively and precisely in units of amount on linear scales having true zero points. Accordingly, although all attempts at prediction based on less than the whole number of factors are imperfect, ultimate perfection of prediction is possible of achievement. The derivation process thus illustrates the general method by which the isolation and use of general and specific factors must be carried forward in similar complex situations before perfect prediction may be achieved.

A second reason for the choice of grade attained as the field of prediction is the need of making clear the complexity of even the simplest and most objective concepts in education. On first thoughts, training would seem to be one of the simplest and most easily measurable of the factors which condition success in school. Actually it is far from a simple factor. The illustration will uncover a number of important problems in this field that are in need of experimental investigation.

For the purposes of the illustration the reader is to imagine that the investigator is an unsophisticated individual, who has come to be interested in the prediction of a child's grade from his age, without realizing that there is more than one factor involved. This assumption of lack of sophistication should not be a difficult one to make: for in many fields of investigation, nay, in most educational fields, all of us are, in fact, extremely unsophisticated. Many still discuss the prediction of school success from scores in intelligence tests as if no other factor than intelligence determined success in school.

The natural procedure in such situations is to express the imagined relationship in mathematical symbols. The basic (assumed) data are

TABLE XLVI
ASSUMED DATA FOR AGE AND GRADE

Age (years)	6	6.5	7	7.5	8	8.5 etc.
Grade (name)	L1	H1	L2	H2	L3	H3 etc.
(symbol)	1	2	3	4	5	6 etc.

in which L and H represent the low and high half grades respectively, and are symbolized by the numbers 1, 2, etc. That is, for convenience each half grade may be considered a unit of training. The figures therefore are numerical symbols for indicating grade attained.

The resulting prediction formula is

$$G = 2(A - 5.5) \text{ or } 2A - 11 \dots \dots \dots A$$

or, in words, the grade of any child may be found by subtracting eleven from twice his age in years.

Let us suppose that in attempting to check the adequacy of the formula, records of three children in school are chosen at random from the school population with the following results:

TABLE XLVII
VALIDATION OF FORMULA BY COMPARISON OF PREDICTED GRADES FROM AGE WITH ACTUAL GRADES ATTAINED

Individuals	Age in Years June 1	Actual Grade	Predicted Grade	Error
A	13	10	15	+5
B	13	15	15	0
C	11	13	11	-2

Suppose that the investigator feels that these results indicate that Formula A is less adequate than is desired. Suppose also that he goes over all his school records and selects every record of boys, ages 10.5 to 13 for which he has complete data from the time the children entered school, and that he obtained 190 such records.¹

Suppose, next, that he computes the age of each child in months to June 1st, 1923, and correlates age in months with grade, obtaining Mean Age = 142.2. Mean Grade = 11.08. S. D. Age = 11.9. S. D. Grade = 2.08. Product moment $r = .74 \pm .02$. A regression equation of the form $Y = C_0 + C_1 \text{ Age}$, or

$$\text{Grade} = -7.46 + .13 \text{ Age} \dots \dots \dots B$$

and a standard error of prediction of 1.39 grades.

If nothing were known about a child except that it was a boy between the ages of 10.5 and 13 years, the best prediction that could be made of his grade would be the average grade of the entire group, or 11.08. The standard error of such a prediction

¹These and the following data and results are the actual data obtained in the investigations upon which this dissertation is based

would be 2.08 grades. The reduction in error effected by basing prediction on age (from 2.08 to 1.39, or .69 units) is a reduction of but 33%. Two thirds of the error as measured by the standard error of estimates based on minimum knowledge is caused by other factors than age.

It should be noted at this point that these "mass" results make certain facts plain, but give no hint of how to proceed to detect the missing factors. Therefore, let us turn to the piecemeal development of a prediction formula for the same situation.

The simplest expression of the relationship between age and grade is that average grade is associated with average age, or using whole units (.08 grades arbitrarily taken as equal to 1.2 months to give round numbers consistent with known facts).

$$\begin{aligned}\text{Average Grade} &= \text{Average Age, or} \\ \text{Grade 11} &= \text{Age 141}\end{aligned}$$

If an individual varies in age from the mean age, he should vary in grade from the mean grade also. The next step, therefore, is to subtract the assumed mean grade from each grade and the assumed mean age from each age, and to investigate the relationship between these differences. (Table XLVIII).

It may be seen by inspection that the relation of grade deviations to age deviations is rectilinear, but extremely variable. On the average, the relationship of age deviations to grade deviations may be expressed as

$$\text{Grade deviations} = C \times \text{Age deviations.}$$

The average value of C was computed,¹ and found to be .13. Therefore the resulting piecemeal formula is

$$G = 11 + .13 (A - 141) \dots\dots\dots C$$

or, in words, the grade of a particular child is equal to the mean grade of the group of children increased by .13 times the amount the age of the child differs from the mean age of the group. This formula if expanded and reduced to simplest form becomes

$$G = -7.3 + .13 A$$

which approximates the conventional formula, B. So far there has been no gain from the piecemeal procedure except that Formula C directs attention to deviations of age from mean age as significant.

However, no critically minded investigator can study Table XLVIII and be satisfied with this formula as an expression of the relationship between age and grade. In view of the scatter in the table, it is clear that the ratio, .13, is merely an average value. Statistically, it is the best number by which to multiply the figures which represent age to secure with the least error² the figures which represent grade, but it includes within itself the effect of all the factors other than age as well as the effect of age. Perhaps this is best made clear by finding the ratio between age and grade deviations for the 190 cases (Table XLIX, A).

At this point it is necessary to define terms. The ratio, .13, is the true value of the age-grade relationship between deviations, only if the true value is defined as the value which represents the entire series of values with the least error. The conception of value upon which prediction is based, however, is that of functional relationship; the idea that if age were the only factor operating, there would be a constant relation between changes in age and changes in grade; that when age operates with other factors, its effect nevertheless may be computed in terms of the constant relationship.

¹See page 178.

²Root mean square error.

TABLE XLVIII

Scatter diagram showing the differences of each child's grade from the mean grade of the group in relation to the differences of same children's ages from the mean age. Mean grade taken as 11. Mean age as 141 months. Number of cases, 190.

Differences from Mean Age	Differences from Mean Grade										Total
	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	
+25							1				1
24											0
23											0
22											0
21							1				1
20							1	3	2		6
19				1	1	1			1		4
18						2	2	1			5
17					1		1			1	3
16							1	1	2	1	4
15					2	1	1	2	2		8
14						2	1	1			4
13					1	1	2	1	2		7
12	1			3		2		1	2		9
11				2			1		2		3
10					1	1	1	2			5
9				1	1	1	1				4
8						1		1	1		3
7						2	1	1			4
6				1		2					3
5				1	1	1	3				6
4				1	1	1	1	1			5
3			2	2	2	3	1				10
2					4						4
1				1	1	1	1				4
0		1	1	1	1	1					5
-1				2	1	1					4
-2							1				1
-3		1		4	2						7
-4				2	1						3
-5		1		1			1				3
-6		1	2	2		2					7
-7				1	1						2
-8			3	1	2						6
-9			2	1							3
-10			1	1							2
-11			3	4			1				8
-12		1	2	1							4
-13			2	1							3
-14	1			1							2
-15		3	1	3							7
-16		2	3	1							6
-17	1	1	1			1					4
-18	1	1	2								4
-19			2								2
-20		3									3
-21			1								1
Total	4	15	28	40	24	27	24	15	12	1	190

Age deviations; Mean +1.2. Median +3.0. Per cent older 54%, younger 43%.

Grade deviations; Mean +.08. Median +0.3. Per cent above average grade 41%, below 46%.

TABLE XLIX

Magnitude of the Variations Caused by the Attenuating Factors as Measured by A, Ratio of Grade Deviations to Age Deviations, True Ratio .167; B, Variations in Age at Entrance from 75 Months, the Mean Age at Entrance; and C, Variations in Failures and Promotions, One Equals a Half Grade Difference from Normal Progress. 190 Cases.

A		B		C	
Ratio	Frequency	Deviation	Frequency	Deviation	Frequency
		25	1		
+ ∞	3	24	0	+3	2
		23	0		
2.0	1	22	1	+2	3
		21	0		
		20	2		
				+1	27
1.0	4	19	0		
.9	0	18	0		
.8	0	17	4	+0	80
.7	1	16	0		
.6	3	15	0		
				-1	45
.5	3	14	1		
.4	4	13	1		
.3	16	12	1		
.2	34	11	4	-2	15
.1	48	10	2		
.0	50			-3	11
		9	1		
		8	1		
-.1	8	7	2	-4	5
-.2	5	6	4		
-.3	1	5	12	-5	2
-.4	4				
-.5	0	4	4		
		3	6		
-.6	0	2	8		
-.7	1	1	13		
-.8	0	0	12		
-.9	0				
-1.0	3	-1	17		
		-2	20		
-∞	1	-3	22		
		-4	18		
		-5	12		
		-6	8		
		-7	5		
		-8	3		
		-9	2		
		-10	1		
		-11	0		
		-12	1		
		-24	1		
Mean*	.13		+.1		-.5
Median.....	.15		-.9		+.2

*Found by exact methods.

A constant relationship, when discovered, is a natural law. Illustration of such relationships in physics are those of temperature to volume and of pressure to volume in gases. An illustration of a complex situation in which several factors determine the changes which take place, and in which constant relationships between various factors enable prediction to be made with certainty is the chemical action of acid on a metal. It is found that there is a single constant relation between the mass of the metal used and the volume of the gas set free, but this relationship is not a series of widely differing ratios. It is a single ratio. Yet if the actual values of the mass of metal used and the volumes of gas set free were treated as the age grade values are treated above, a series of ratios similar to those in Table XLIX, A is obtained.³

Appreciation of these facts would lead an investigator, studying Table XLVIII, to ask himself,

What is the constant value of the relationship between age and grade?

What are the other factors acting whose effects result in the distorted values of Table XLIX, A?

The investigator started with the common sense observation that normal progress in school would make the ratio between age and grade deviations .167; that is, normally it takes six months of time (five months of schooling and one month vacation) to carry a child through one grade. If this value is inserted in Formula C above, it becomes

$$G = 11 + .167 (A - 141) \dots \dots \dots D$$

or expressed in conventional form

$$G = -12.5 + .167A \dots \dots \dots E$$

For children whose school progress was affected by no factors other than age, Formula E would yield absolutely perfect prediction. The differences between Formulae C and E constitute a measure of the degree to which assumptions implicit in the use of the regression technique fail to fit this situation. In terms of the problems previously discussed, the statistical procedure used in the derivation of Formulae B and C is not valid.⁴

The piecemeal procedure thus stimulates one to question the validity of the statistical methods used to isolate the factors and to prove that Formula E modified to allow for the action of the other factors yields perfect prediction in all cases.

The first step in analysis would be to see whether any cases of ratio, .167 occur among the 190. Actually nine were found, but a scrutiny of the records of these children revealed the fact that not one of them had progressed through the school at the rate of two half-grades a year. One child had entered early and failed of promotion at some point; another had entered late and received double promotion; and so on. The value .167 was found in each case to be the resultant effect of the play of several factors, and not the value of the age-grade relationship as expected. These analyses, however, brought to light the fact that two of the factors operating are the age at which a child enters school, and whether he is double promoted or failed at any point in his career. These factors will be called the entrance and promotion factors respectively. The entrance factor is defined as the number of months earlier or later than 75 months (the average age at entrance of this group of boys) that a child enters school. The promotion factor will be defined as the number of grades of failure or irregular promotion which a child receives.

³The ratios of observed volume in cubic centimeters to mass (x 100) in grams for the 29 cases (page 173) are:

Ratio	Frequency	Ratio	Frequency	Ratio	Frequency	Ratio	Frequency
22.0	2	13.0	4	11.0	5	9.0	6
		12.0	0	10.0	9	8.0	3

Mean value 11.42. True value under standard conditions 9.20.

⁴See page 173.

Both of these are independent variables. They have no fixed relation to either age or grade. The month of a child's birth, the views of a child's parents in regard to the proper age to enter school, the health of the child, the family situation in regard to travel, social position, etc., are a few of the many variables which determine age at entrance. In a general way the compulsory attendance law and the rigor of its enforcement must also be included. The extent to which the entrance factor attenuates the relationship between age and grade is shown by a tabulation stimulated by the piecemeal procedure (Table XLIX, B).

In similar fashion, promotion factor is an independent variable. Children may win extra promotion, or are failed, in all grades and at all ages. Also, the factors which determine failure and promotion are many. Intelligence is the chief instructional factor,⁵ but available seats, physical size, personal standing with teacher are a few of the other factors which readily occur to anyone familiar with conventional school practice. The attenuating influence of these factors is shown by a tabulation of the value of the promotion factor for the 190 children being considered (Table XLIX, C).

The situation may be summarized by saying that out of 190 boys between the ages of 10.5 and 13, only 12, or 6%, entered at the average age, and only 80, or 42%, made normal progress. Perhaps it is still more striking to say that only five children were found who had both entered at the average age of entrance and made normal progress. Moreover a study of Table XLIX B and C will show that the distributions are both irregular and skewed. Children tend to fail more than to be double promoted, while their ages of entrance tend to be lower than 75 rather than higher. These variations from the conditions assumed to exist by the conventional regression equation technique account for part of the failure of the regression equation to express the true relation between age and grade.

The fact that five children were found who entered at the average age and made normal progress suggests that these cases might be used to verify Formula E. However, the ratios in the four cases proved to be .150, .200, .333, .333, .333. To make these ratios come out .167, as called for by the theory of normal progress, it was necessary to add two months to one of the ages, and three months each to the other four. A study of the differences between the two types of cases ultimately led to the discovery that one of the children entered school in February, while the other four entered in September. Following up this clue, it was soon apparent that in every one of the 190 cases perfect prediction could not be made, even allowing for the factors so far considered, without making a correction of two or three months, depending upon whether the time of entrance was February or September.

Once the fact was discovered, the reason was not hard to find. Grades are computed as if completed. That is, a child in the sixth grade on the first of June was counted as having finished the grade, but his age was computed exactly in months to the first of June. The correction operates to make age and grade comparable. June and one month of vacation (July) make up the two months added to age for those who enter in February, while three for those entering in September includes also the August month of vacation which precedes September.

The foregoing may be summarized by saying that the raw data of ages and grades are, statistically, both incomparable and invalid as measures of the quantities to be related. Before the true relation between age and grade can be determined, corrections must be applied as follows:

1. Ages of children must be made comparable to grade by adding two months in case of February entrances, and three months in the case of September entrances.

⁵Product moment coefficient of correlation between intelligence scores expressed in standard scores for age (variability units) and deviations from normal promotion, $r = .44 \pm .04$.

2. Ages of children must be increased or decreased by as many months as they were late or early in entering school.
3. Ages of children must also be increased or decreased by six months for each time they were failed or given an extra unit of promotion.

That is,

$$EA = AA + C + E + 6P \dots\dots\dots F$$

in which EA equals age equated with grade, AA equals actual age at June 1st, C the equating factor of 2 or 3 for month of entrance, E the equating factor for variation in age at entrance, and P the equating factor for variation in promotion. Note that these factors are all of the type designated as specific. They operate on individuals in an individual manner so that no general statement of a law of action is possible. If it was desired to resolve these corrections into statements of the action of general factors, it would be necessary to isolate the factors which cause late and early entrance and those which cause promotion. That, however, is not the province of the present study.

If now the factors affecting age are written into Formula E, it becomes

$$\begin{aligned} G &= \text{Mean } G + C_1 [(A + C + E + 6P) - \text{Mean } A] \text{ or} \\ G &= 11 + A/6 + C/6 + E/6 + P - 141/6, \text{ or} \\ G &= -12.5 + .167 (A + C + E) + P \dots\dots\dots G \end{aligned}$$

an equation which yields perfect prediction for the 190 cases.

If the investigator were a sophisticated individual, who after careful study of the situation had isolated all these factors by general observation, and had computed ages at August 1st so that he had at the outset as his basic data, age at August 1st, age at entrance, and grades promoted or failed, the regression equation of the form

$$Y = C_0 + C_1 X_1 + C_2 X_2 + C_3 X_3$$

would be

$$\begin{aligned} \text{Grade} &= 0 + .167 \text{ age} - .167 \text{ age at entrance} + \text{promotion, or} \\ &= .167 (A - E) + P \dots\dots\dots H \end{aligned}$$

which throws a great light on the whole situation. For it makes clear an important fact which a truly sophisticated individual, planning on using a method of analysis, in a scientific fashion, would have ascertained at the outset; namely, that age does not determine grade, so that to apply the correlation technique to the age-grade data is to obtain a coefficient which expresses merely a type of defined mathematical relationship between two series of numerical terms. It is not age that determines grade, but "passage of time" as is clearly indicated in Formula H, but concealed in Formula A to G. The conclusion to be drawn is that the correlation technique yields measures of relationship which are true to reality only when all the factors are considered, and all the data conform to the basic assumptions. The data for general factors must be corrected for the action of specific factors before the general relationship may be reliably determined. In other words the law of the single variable holds in statistics as elsewhere.

A precisely similar situation arises when grade or school marks are correlated with intelligence scores. Scores in a group intelligence test represent a complex which is a function of at least capacity, maturity, and quite probably training as well. There is no functional relationship between promotion and intelligence as represented in the conventional scores for these two qualities, but if each could be freed by correction from attenuating factors there would probably be a constant functional relationship between a pure measure of capacity and the number of extra failures or promotions.

It is the discovery of such functional relationships which is the goal of the investigator. They are never revealed by the correlation technique until as in the case of Formula H above all the factors operating are accounted for and the standard error of estimate falls to zero. The piecemeal method with constant investigation of divergent cases enables the research agent to note the character of the changes which take place as one factor after another is introduced into the prediction formula. It thus operates as a powerful stimulus to analytic thinking.

As an illustration of the need for careful experimental or analytical determination of relationships between factors before the correlation technique is employed, suppose the investigator considers age and age at entrance to be the two factors which determine grade and computes the correlations. He would obtain for the 190 cases

TABLE L
CORRELATION BETWEEN GRADE, AGE, AND AGE AT ENTRANCE

	Grade	Age	Age at Entrance
1. Grade.....	..	.75	— .06
2. Age.....	.75	..	.13
3. Age at Entrance.....	— .06	.13	..

The partials are

r , grade — age . age of entrance constant = .7610

r , grade — age at entrance . age constant = —.2421

R , grade . age + age at entrance = .7622.

in spite of the fact that

r grade . passage of time corrected for promotion = 1.00.

Analysis alone can save an investigator from being misled by results like these, which, while statistically true, are meaningless as measures of the type of relationship which it is desired to find.

The illustration was also used to call attention to another series of problems, namely, the need for the definition of units of training. By Table XLIX, B the boys in this group are proved to vary in age of entrance from —24 to +25 months from 75 months. But capacity matures with age. Consequently first grade training given to a boy 55 months old should produce quite different effects from those secured when identically the same training is applied to a boy 99 months old. It may prove possible to use a constant value for the first grade training and apply a correction factor for maturity. But two more difficult problems in measuring amount of training arise when children are failed or given extra promotions. If a child is in school six semesters and passes through three half-grades, has he had six units of training, or three, or some intermediate number? Similarly, if in six semesters in school he passes through twelve half-grades, how many units of training has he received? Certainly Table XLXI, C makes it plain that before units of training can be used in prediction formulae, answers to such questions as these must be found by experimental solution under the law of the single variable. Probably the absolute unit of training will ultimately be defined as that amount of training of a given type which causes a change of one unit of absolute ability when acting on one unit of absolute capacity during a unit period of time. At present, however, such a definition serves merely as a means of emphasizing a series of important research problems.

INDEX

- Ability
 Adding, 85, 87, 108, 221
 Definition, 46, 49, 50, 82f
 General and specific abilities, 82-88 187, 257
 Handwriting, 98, 110ff, 133, 197ff
 Measurement, indirect, 82
 Reading, 88ff, 143-144
 Relation to performance, 48
 Spelling, 140-143
 Variation, 14
- Accomplishment quotient, 25
- Acknowledgments, 3
- Adding, 85, 87, 108, 135, 221
- Adjustment factor
 Angle of slant of adjustment line, 121ff
 Composition, 137
 Derivation of formula, 245
 Handwriting, 118ff
 Lines of adjustment, 121ff
 Reading, 144ff
 Spelling, 141ff
 Stanford Achievement Tests, 144ff
 Typewriting, 137
- Age
 Factor in prediction of Stanford score, 192ff
 Factor in prediction of handwriting score, 197
 Relationships
 Age-dash, age-height, age-lung, age-target-throw, 91ff
 Age-grade, 56, 251-259
 Age-grade-progress, 56
 Age-height-weight, 175ff
- Aims of dissertation
 General statement, 17
 Specific statement, 19
- Alchemy, 85
- Anthropometric laboratory, 24
- Appraisal
 Performance scales, 140ff
 Results in educational measurements, 26
 Worth of adopted standard of success, 96
- Arithmetic
 Adding, 85, 87, 108, 136, 221
 Individual variation, 136
- Army Intelligence Tests, Detroit form
 Use in investigation, 54, 56, 208
 Factor in prediction
 Stanford score, 192ff
 Handwriting score, 197
 Sample, 217
- Ashbaugh, E. J., 143
- Assumptions, fundamental, 82
- Bagley, W. C., Educational determinism, 31
- Baldwin, B. T., 24, 180, 181
- Bases of record of behavior, 102
- Behavior
 Aspects in rate tests, 106
 Creative resultant, 32f
 Possibility of analysis, 86
 Relation to ability, 82
- Bergman, W. G., 4
- Bernard, Claude, 22
- Binet, A.
 Concept of mental age, 25
 Laboratory of Physiological Psychology, 25
 Scale for measuring intelligence, 25
- Bird, Charles, 165
- Birkhoff, G. D., Development of concepts, 34
- Blackstone, E. G., 4, 137
- Boas, Franz, 24
- Bowditch, H. P., Measurement of physical growth of children, 24
- Boyle, R., 16
- Brown, H. A., Reading ability, 113
- Brown, W., 83, 93, 94
- Build
 Computation of build factor, 181ff
 Concepts of build, 181ff
 Relation of hereditary factors, 199
- Bulletin, Board of Education (1923), 12
- Caldwell, O. W., 12
- Calkins, M. W., Problem of reality, 30
- Capacity
 Definition, 42ff
 Relation to training, 14f
- Carver, H. C., 4, 247
- Cattell, J. McKeen
 Pioneer in experimental investigation, 23
 Definition of heredity, 38
- Cavendish, H., 16
- Change, 80f
- Characteristics of children, 37f
- Chemistry, 85, 173ff
- Cody, Frank, 3
- Commonwealth Fund, 3

- Comparison
 Average school mark and various factors, 63-78
 Development in physical sciences and science of education, 16
 Effects of scarcity of food upon children of different ages, 170ff
 Effects of two stimuli on effort, 125ff
 Relative complexity, 93
 Variabilities of rate and performance tests, 157-160
 Variations in immediate and delayed re-testing, 100
- Complexity and variability, 91
- Composition, 136f
- Conclusions
 Analytical prediction, 200
 Analytical scientific method in education, 94
 Subjective and objective measures of success, 78
 Variability of scores on performance scales and rate tests, 160
 Variability of retests, 101f
- Conditions of investigation, 55
- Constancy of scores in performance scales, 157f
- Constants, values in prediction formulae, 248
- Construction of performance scales, 140
- Conventional prediction formula
 Compared to piecemeal formula, 175ff, 251-259
 Proof of equivalence to piecemeal formula, 247-248
- Control mechanisms, 47, 49
- Control, regions of
 Definition, 108
 Intelligence tests, 147
 Reading tests, 143-147
 Spelling tests, 141-143
- Cook, W. A., 143
- Cooley, C. H.
 Dangers of formalism, 31
 Heredity and environment, 43
- Correction of raw scores, 139
- Correlation, 91ff, 158, 173ff
- Curtis, S. A., 12, 13, 88, 90, 106, 108, 124
- Criticisms of prediction procedure, 184ff, 195-196
- Cubberly, E. P., 11
- Dalton, John, 16
- Dangers
 Application of factor theory, 31
 Conventional interpretation of test scores, 138
- Data of investigation
 Types, 53
 Conditions under which secured, 55ff
- Davenport, C. B., 181
- Davidson, P. E., 88
- Dearborn, W. F., 60
- Defects of prediction procedure, 184ff
- Definition
 Ability, 46
 Capacity, 43ff
 Conditions of action of individual, 49
 Environmental forces, 44
 Factor concept, 30ff, 35
 Heredity, 38 ff
 Maturity, 44f
 Natural law, 29
 Need of definition, 89
 Performance, 50
 Success, 59
 Terms and concepts used, 37-52
 Training, 44
- Derivation of formulæ for effort and adjustment, 243-245
- Desmoulins, L., 22
- Determination of the effect of training upon intelligence scores, 172
- Determinism, educational, 30f
- Detroit
 Addition Test, 208, 221
 Army Intelligence Test, 54, 56, 208, 217
 Reading Tests, 208, 219
 Superintendent of Schools, 3
 Teachers College, 3
 Training Schools, 3, 55
- Dewey, John, 12
- Difficulty
 Nature of problem of difficulty, 105
 Problem for research, 202
 Rate, quality and difficulty in performance, 102
 Rate tests, 106ff
 Solutions of problem of difficulty, 105f
- Dvorak, A., Decease of I. Q. with age, 196
- Education
 American ideal, 11
 Present and future significance, 11
 Present inefficiency, 12-16
 Present need, 16
- Educational age, 25
- Educational determinism, 31
- Educational quotient, 25
- Effect of effort on handwriting scores, 126
- Effect of objective criterion, 75-78

- Effects not additive, 32
- Effort factor
- Derivation of formula, 243,-244
 - Direct measurement, 132-135
 - Effect upon adjustment lines, 130
 - Evidence of effect on scores
 - Handwriting, 124ff
 - Composition, 137
 - Spelling, 140ff
 - Typewriting, 137
 - Law, 131
 - Problem for research, 202
- Elliott, C. M., 98
- Emerson, William R. P., Relation of height, weight and malnutrition, 187ff
- Emotional survey
- Place in program of investigation, 54, 56
 - Instructions for rating, 225
 - Sample record card, 229
- Emotional mechanism, 47, 49
- Environment and heredity, 40ff
- Environmental forces, 44
- Eugenics Laboratory, 24
- Evidence of inefficiency in public education, 12-16
- Experience, 44
- Factor
- Factor concept
 - Dangers of application, 31
 - Definition, 29, 35f
 - Difficulties of application, 33ff, 89
 - Genesis, 30
 - Nature and importance, 29-36
 - Possibility of application, 194, 201
 - Purposes, 36
 - Relation to reliability, 98
 - Factors affecting teacher judgment, 63ff
 - General and specific factor theory, 83, 183, 257
 - Major factors determining success, 50, 79
 - Types of factors contributing to conclusions in educational measurements, 50
- Farrand, Max, 3
- Fechner, Gustav, 22
- Field of variation, 131ff. See *Variation*
- Final Report, Committee of School Inquiry, 13
- Flexnor, Abraham,
- Formalism, educational, 31
- Formula
- Ability handwriting, 116, 133
 - Adjustment factor, 131, 245
 - Adjustment line, 121
 - Age-grade, 252, 254, 256
 - Angle of adjustment, 121
- Formula
- Effort factor, 131, 243-244
 - Gates', handwriting ability, 116
 - Grades and various factors, 252ff
 - Laws of gases, 174
 - Prediction of weight, 175ff
 - Prediction of Stanford score, 191ff
 - Prediction of handwriting score, 196ff
 - Proof of equivalence of conventional and piecemeal prediction formulæ, 247-248
- Franzen, R., 25
- Galton, Sir Francis
- Experimental psychologist, 22
 - Heredity and environment, 43, 204
- Gary survey 3, 106
- Gates, A. I.
- Handwriting ability, 115ff
 - Relation of physical and mental development, 186
 - Specific maturities 83, 94
- General and specific factor theory, 83-88, 186, 194, 197, 199, 257ff
- Generalizations, 201
- Genesis of factor concept, 30
- Goddard, H. H., 25
- Grade-age relationship, 56, 251-259
- Growth
- Hereditary factor and growth, 180ff
 - Mental growth, 189f
 - Phases of growth, 82
 - Physical growth, 24, 180, 181
- Hall, M., 22
- Hamilton, Sir W., 22
- Handwriting
- Adjustment factor, 118ff, 245
 - Computed scores, 115-118
 - Effort factor, 243-244
 - Individual variation, 110ff, 123ff
 - Law of ability, 133
 - Prediction of scores, 196-199
 - Sample handwriting test, 221
 - Scoring, 113
 - Testing conditions, 112-113
 - Variation in computed score, 117
- Hart, H., 196
- Hastings, W., 24
- Health, relation to weight, 181ff
- Height
- Height-weight-age relationship, prediction from, 175ff
 - Height, weight, and malnutrition, 187
 - Height, weight variations, 98, 169ff
 - Reliability of height and weight measurements, 98

- Helmholtz, H. L., 22
- Heredity
 Genetic process determining heredity, 39ff
 Hereditary capacity, 43ff
 Hereditary differences in maturation of intelligence, 189
 Heredity and environment, 40ff
 Meaning of term, 38-44, 49, 50, 162
 Measurement of hereditary factor, 161ff
 Relation between hereditary factors, 199
 Relation to height, 180
- Herring, J. P., 196
- Hillegas, M. B., 104
- Hines, H. C., Stanford tests as performance scales, 61
- Historical background of investigation, 21-27
- Hollingsworth, L. S., 82
- Hudleson, E., 136
- Hughes, W. H., 71
- Huntington, E. V., 176
- Illustration of increase of reliability, 99
- Illustration of use of prediction formulæ in study of individuals, 188, 251-259
- Implications of study, 203
- Index of build, 181ff
- Individual differences
 Causes, 43f
 Capacity, 25
 Heredity, 38
- Individual variation
 Performance, 110ff, 123ff
 Weight, 181
- Industrial revolution, 16
- Inefficiency of public education, 12-16
- Intelligence
 Effect of training upon scores, 178
 Factor in prediction of Stanford and hand-writing scores, 192ff
 Group intelligence tests, 25
 Individual intelligence tests, 24
 Measurement, 24, 25, 147ff
 Mental age, 189
 Quotient, 25, 31, 190, 195, 199
 Relation to ability in reading, 14f
 Relation to success, 14
 Tests given, 53f
 Utility of intelligence tests, 148
- Interpretations
 Data supporting objective measure of success adopted, 78
 Educational situation, 15
- Instructions for giving and scoring certain tests, 231-241
- James, William
 Description of first stage of consciousness, 30
 Estimate of progress in educational method, 26
- Jennings, H. S., 39ff
- Judd, C. H., 3
- Justification of adoption of narrow definition of success, 59-78
- Kant, I., 30
- Kaufman, H. J., 4
- Kelley, T. L.
 Statistical method, 57, 247
 Value of school marks, 60
- Laboratory of Physiological Psychology, 25
- Lavoisier, A. L., 16
- Law
 Adjustment, 134
 Effort, 134
 Gases, 174
 Laws of variation, 133
 Natural law, 29
 Single variable, 29, 172
- McCall, W. A.
 Reliability, 97, 99
 Scale construction, 140
- McFarland, Ernest, W., 3
- Madsen, I. N., 196
- Malay, Theresa, 4
- Manny, F. A., 177, 187
- Marks, comparison of average school mark and various factors, 63-78
- Maturity
 Definition, 44-45, 49, 50, 162
 Effect upon weight and height, 190-191
 Effect upon scores, 194, 197
 Hereditary differences in maturation of intelligence, 189
 Matured capacity, 167, 197
 Maturation of pecking instinct, 165-167
 Measurement, 161ff
- Measurement
 Activities, 53-57
 Ability, 82
 Education, 25f, 34
 Heredity, maturity, and training, 161-200
 Quotient, 102-105
 Intelligence, 24-25, 54f, 208
 Performance, 79-160
 Physical characteristics, 23, 24
 Rate, 102
 Success, 60ff
- Mechanisms, 46ff

Mental

Age, 189

Mechanisms, 46ff

Mental growth curves, 189

Methods, statistical, 113ff, 202

Miller, W. S., Reliability of I. Q., 196

Moeblman, Arthur B., 4

Moore, E. C., 82

Monroe, P., 22, 35

Monroe, W.

Scale construction, 140

Teachers' marks, 60

Motor mechanism, 47, 49

Motor mental tests, 53ff, 197, 222

Multiple correlation, 172, 192, 194

Murdoch, Catherine, 104

National Intelligence TestFactor in prediction of handwriting and
Stanford scores, 192ff

Sample test, 215

Use in investigations, 54, 56

Natural law, 29

Nature and nurture, 204

Norsworthy, N., 24

Objections to measure of success used, 61

Objective criterion, effect of, 75-78

O'Shea, M. A., Variation in spelling perform-
ance, 143

Otis, A.

Logical aspects of Binet's scale, 148

Statistical method, 143

Test validity, 88

Test reliability, 99, 143

Pearson, Karl, 24

Performance

Conditioning factors, 47f

Definition, 47, 50

Measurement, 79-160

Misconceptions in measurement, 109f

Relation to ability, 48, 50

Relation to reliability, 98

Scales, 139ff

Periods of evolution of psychology, 21

Physical characteristics, measurement, 23, 24

Piecemeal derivation procedure

Criticism, 184ff

Development, 177ff

Equivalence to conventional formula, 247-
248

Illustration of use, 251-259

Prediction of weight, 176ff

Prediction of Stanford Achievement Scores,
191ff

Uses, 186ff

Porter, W., 24

PredictionAccuracy of prediction of Stanford scores,
193

Correlation technique, 175ff

Criterion for judgment of truth, 49

Goal of prediction, 16

Illustration, conventional and piecemeal
prediction procedure, 251-259

Piecemeal derivation, 178ff

Prediction of

Handwriting, 196-199

Success, 17ff

Stanford scores, 191-194

Weight, 175-184

Proof of equivalence of piecemeal and con-
ventional formulæ, 247-248

Present need in education, 16

Priestly, Joseph, 16

Pritchett, H. S., Criticism of public school edu-
cation, 12**Problems**

Direct measurement of effort, 133ff

Measurement of difficulty, 105-132

Measurement of quality, 102-105

Measurement of reading ability, 144

Piecemeal derivation procedure, 184f

Problems raised by investigation, 202

Reliability, 97-102

Test validity, 88-94

Utility, 94-97

"Why Children Succeed"

General answer, 194

General outline, 16f

Nature of problem, 79-82

Products yielded by investigation, 201ff

Program of investigation, 18, 53ff

Promotion, relation to grade attained, 257-259

Proof of variation of performance, 110, 150ff

Psychology

Descriptive, 21

Educational, 23

Experimental, 22

Periods, 21-23

Psycho-physicists, 22

Purposes of analyses of measurement, 161

Quality

Bases of record, 102

Measurement, 102-105

Methods of treatment of rate and quality
scores, 113ff

Nature of problem of quality, 102

Problem for research, 202

Relation to rate and difficulty, 102

Solutions of quality problem, 103ff

- Quotient
 Accomplishment, 25
 Intelligence, 25, 31, 190, 195, 199
- Range of problems studied, 16-17
- Rankin, Paul T., 4, 137
- Rasey, Marie I., 3, 140
- Rate
 Behavior in rate tests, 106
 Difficulty, quality, rate, and performance, 102
 Measurement by rate tests, 107
 Methods of treatment of rate and quality scores, 113ff
 Rate tests, 106ff
 Timing of rate tests, 107
 Types of performance on rate tests, 108
- Reading
 Ability, 88ff
 Difficulty scales, 143-144
 Sample reading tests, 218-220
 Thorndike tests, 13, 143
 Variation of performance, 144-148, 150-156
- Reasons for choice of tests, 207-209
- Record cards, social and emotional surveys, 229
- Records and measurements on file, 56
- Relation
 Age and grade of children tested, 56
 Age and various physical traits, 91ff
 Correlation, complexity, and variability, 191ff
 Factors contributing to conclusions, 50
 Height, weight, age, 175ff
 Height, weight, and malnutrition, 187
 Hereditary factors, 199
 Intelligence and success, 14, 192ff
 Maturity of thought and motor processes, 163
 Promotion and success, 74ff
 Teacher estimate and success, 62ff
 Testing conditions and performance, 112f
 True volume, observed volume, temperature, and pressure, 174
 Various factors and grade status, 66ff
 Various factors and objective measures of success, 75ff
 Various factors and school marks, 64ff
- Reliability
 Explanations of variations in retests, 98f
 Illustrations of increasing reliability, 99
 Nature of problem of reliability, 97f
 Problem for research, 202
 Reliability of I. Q., 195-196
 Scoring handwriting, 114
- Research departments, 3, 4
- Results in tests under varied conditions, 113ff
- Retest and reliability, 97ff
- Rice, J. M., Experimental investigation, 12, 23
- Richards, A. M., 88
- Rietz, H. L., 176
- Riley, J. L., 12
- Ruch, G. M., 61, 97
- Rugg, H. O., 57
- Sample tests and records, 212-229
- School marks
 Relation to
 Maturity, 64ff
 Sex, 67ff
 Stanford scores, 67ff
 Teachers estimate of success, 63f
 Teachers estimate of rank in class, 64ff
 Training, 64ff
 Difficulty in using, 62ff
- Scheele, K. W., 16
- Scientific method, 16, 22
- Scoring and tabulating conditions, 113
- Silent reading, 13
- Single variable, law of, 29f, 81, 172
- Smith, E. R., 12
- Social factor, relation to Stanford score, 194
- Sociological survey, 3, 53ff, 223, 229
- Solutions of difficulty problem, 105f
- Solutions of quality problem, 103ff
- Spearman, C.
 Contributor to experimental method, 24
 Estimate of progress in method, 26
 General and specific factor theory, 83
- Specific factor theory, 82-88, 186, 194, 257
- Spelling, 53f, 140ff, 233-234
- Standard error of estimate, 91, 175ff, 252f
- Stanford Achievement Tests
 Difficulty scale, 144-147
 Fields of control illustrated, 147
 Measure of success,
 Prediction of score, 191ff
 Relation of score to various factors, 67ff, 192ff
 Sample test, 212
 Variation of scores in immediate and delayed retesting, 100
- Starch, Daniel
 Ability score in handwriting, 115
 Unreliability of teachers' marks, 60
- Statistical procedure, 173ff
- Stecher, L. I., 189f
- Stern, W., 25
- Stone, C. W., 13

- Success**
 Definition, 18f, 59-60
 Measurement, 60ff
 Relation of various factors to success, 63ff
 Teachers' estimate of success, 63ff
Summary of definitions, 49
Summary and generalizations, 201
Sunday school attendance, 194
Survey
 Gary, 3, 106
 Salt Lake City, 13
 Virginia, 13ff
Teacher judgment, 62ff
Terman, L. M., 25, 61, 147, 148
Terms, definitions of, 37-51, 201
Test
 Adding, 53, 85, 87, 108, 136, 221
 Binet, 25, 147
 Conditions of testing, 112-113
 Intelligence, 25, 53f, 147, 208, 215
 Motor-mental, 163f
 Reading, 13, 53, 143, 218-220
 Reasons for choice of tests, 60ff, 207f
 Scores available, 53
 Stanford. *See Stanford*
 Terman, 25
 Thorndike-McCall, 89, 144ff, 208, 218
Thesis
 Aims, 17, 19
 Historical background, 21ff
 Limitations, 19
Thiele, C. Louis, 148
Thomson, G. H., 83, 93, 94
Thorndike, E. L.
 Classification of intelligences, 46
 Criticism of scientific method in education, 32
 Definition of maturity, 45
 Factor concept, 35
 Handwriting score, 115f
 Heredity and environment, 43
 Place in growth of educational science, 23
 Quality problem, 104
 Reading ability, 143
 Specific abilities, 83ff
 Thorndike-McCall reading tests, 13, 32, 89, 208, 218
 Value of transfer, 96
Tirrell, H. A., 12
Trabue, M. R., 110
Training
 Definition, 44, 49, 50, 162
 Factor in prediction of scores, 192ff
 Measurement, 161ff
Truth, 49
Types of performance on tests, 108ff
 Validity determinations, 90
Typewriting, 137f
Unpublished studies
 E. G. Blackstone, 137
 P. T. Rankin, 137
 Marie I. Rasey, 140
 C. L. Thiele, 148
Uses of piecemeal derivation procedure, 186ff
Utility
 Definition, 95
 Illustration, 96f
 Problem for future research, 202
Validity of tests
 Definition, 88
 Difficulties of determination, 89f
 Justification of theory adapted, 90ff
 Nature of problem, 88-89
 Types of determinations, 90
Validity of statistical procedure
 Illustration of problem, 173ff
 Nature of problem, 173
 Prediction formulae, 175ff
 Problem for research, 202
Variability
 Comparison of variability of scores on tests, 157-160
 Computed handwriting scores, 117
 Performance, 110, 120ff, 136ff, 140, 143ff
 Relation to complexity, 91ff
Virginia Survey, 13
Ward, H. T., 12
Weber, W. E., 22
Weight-height relationship, 98, 169ff
Weight-height-age relationship, 175f
Weight, height and malnutrition, 187
Whipple, G. M.
 Definition of maturity, 45
 Experimental investigator, 23
 Formula for efficiency index in handwriting, 115
Whitney, Dean A. S., 3
Williams, H. S., 16, 22
Witmer, L., 24
Wood, T., 24
Woodrow, H., 84
Woody, Clifford, 4
Wundt, W.
 Experimental psychologist, 22
 Principle of creative resultants, 32
Yerkes, R. M., 25
Yoakum, C. S., 25

LIST OF AUTHORS AND TITLES TO WHICH REFERENCES ARE MADE IN THE TEXT

(The figures in parentheses indicate pages in the text on which the references appear.)

-
- | | |
|---------------------------------------|---|
| Ashbaugh, E. J. | "Variability of Children in Spelling." <i>School and Society</i> , January 18, 1919. (143) |
| Bagley, W. C. | "Educational Determinism." <i>School and Society</i> , 1922, 15:373, April 8. (31) |
| Baldwin, B. T. | "Physical Growth of Children from Birth to Maturity." <i>University of Iowa Studies in Child Welfare</i> , No. 1, 1921. (24, 180, 181) |
| —————, and
Stetcher, L. I. | "Mental Growth Curves of Normal and Superior Children." <i>University of Iowa Studies</i> , 1922, Vol. II, No. 1. (189, 190, 195) |
| Binet, A. | Many articles in <i>L'Annee Psychologique</i> . 1895-1911. (25) |
| Bird, Charles | "The Relative Importance of Maturation and Habit in the Development of an Instinct." <i>Pedagogical Seminary</i> , 1925. (165) |
| Birkhoff, G. D. | "The Origin, Nature and Influence of Relativity." <i>The Scientific Monthly</i> , April, 1924, XVIII:225. (34) |
| Blackstone, E. G. | "The Measurement of Typewriting Ability," an unpublished study. (137) |
| Bowditch, H. P. | "The Growth of Children." <i>Tenth Annual Report</i> , 1879, <i>State Board of Health</i> , Mass. (24) |
| Brown H. A. | "The Measurement of the Efficiency of Instruction in Reading." <i>Elementary School Teacher</i> , June, 1914. (113) |
| Brown, W., and
Thomson, G. H. | <i>The Essentials of Mental Measurements</i> . 1921. (83, 93, 94) |
| Caldwell, O. W. and
Courtis, S. A. | <i>Then and Now in Education</i> . 1923. (12) |
| Calkins, M. W. | <i>The Persistent Problems of Philosophy</i> . 1907. (30) |
| Cattell, J. McKeen | "Mental Tests and Measurements." <i>Mind</i> , Vol. XV, 1896. (23)
"An Interpretation of Intelligence Tests." <i>The Scientific Monthly</i> , May, 1924, XVIII:509. (38) |
| Cook, W. A. and
O'Shea, M. V. | <i>The Child and His Spelling</i> . 1914. (143) |
| Cooley, C. H. | <i>The Social Process</i> . 1922. (12, 31, 43) |
| Courtis, S. A. | "The Construction of Measuring Instruments in the Field of Education." <i>Scientific Monthly</i> , Sept., 1925. (106)
<i>Measurement of Classroom Products</i> . 1919. (13, 88, 90, 106, 124)
<i>Final Report, Committee of School Inquiry</i> . (13, 106)
<i>Then and Now in Education</i> . 1913. (12, 13) |
| —————, and
Caldwell, O. W. | |
| —————, and
Rankin, P. T. | <i>The Measurement of Composition Ability</i> , an unpublished study. (186) |
| —————, and
Thorndike, E. L. | "Correction Formulæ for Addition Tests." <i>Teachers College Record</i> , Jan., 1923. (88, 90, 108, 124) |

- Cubberly, E. P. Public School Administration. 1916. (11)
- Davenport, C. B. Body Build and Its Inheritance. 1923. (181)
- Davidson, P. E. and Richards, A. M. "Correlations of Single Measures of Some Representative Reading Tests." School and Society, September, 1916. (88)
- Dearborn, W. F. "The Relative Standing of Pupils in the High School and the University." Bulletin No. 312, University of Wisconsin, 1909. (60)
- Dewey, John Democracy and Education. 1920. (12)
- Dvorak, A. "The Relation of I. Q. to the Prognosis of Special Class Pupils." School and Society, June, 1924. (196)
- Elliott, C. M. "Reliability of Height and Weight Measurements." Detroit Educational Bulletin, June, 1921. (98)
- Emerson, W. R. P. and Manny, F. A. "The New Weight-Height Tables and Malnutrition." Archives of Pediatrics, Oct., 1924. (177, 187)
- Franzen, R. "The Accomplishment Quotient." Teachers College Record, Vol. XXI, 1920. (25)
- Galton, Sir Francis Hereditary Genius. 1869. (22, 23)
- Inquiries into Human Faculties. 1883. (43)
- Gates, A. I. "The Educational Significance of Physical Status and Physiological, Mental, Emotional, Social Maturity." Teachers College Record, Vol. XXV, No. 3, May, 1924. (186)
- "The Nature and Educational Significance of Physical Status and of Mental, Physiological, Social and Emotional Maturity." Journal of Educational Psychology, XV, No. 6, Sept., 1924. (83, 186)
- "The Relation of Quality and Speed of Performance." Journal of Educational Psychology, March, 1924. (116)
- "Study of Comprehension in Reading by Means of a Practice Experiment." Journal of Educational Research, 1923. (88)
- Goddard, H. H. "The Binet-Simon Measuring Scale for Intelligence." Training School Bulletin, January, 1910. (25)
- Hamilton, Sir W. Discussion on Philosophy, Education, and University Reform. Monroe's Cyclopedia of Education. (22)
- Hart, H. "The Slowing Up of Growth in Mental-Test Ability." School and Society, November, 1924. (196)
- Herring, J. P. "Avery's Comparison of Stanford and Herring Revisions." Journal of Educational Psychology, September, 1924. (196)
- Hillegas, M. B. "A Scale for the Measurement of Quality in English Composition by Young People. Teachers College Record, September, 1912. (104)
- Hines, H. C. A Guide to Educational Measurement. 1923. (61)
- Hollingsworth, L. S. Special Talents and Defects. 1923. (82)
- Hudleson E. "English Composition." Twenty-Second Yearbook, Part I, National Society for the Study of Education. 1923. (136)
- Hughes, W. H. "Organized Personnel Research and Its Bearing on High School Problems." Journal of Educational Research, Dec., 1924. (71)

- Huntington, E. V. *Curve-Fitting by the Method of Least Squares and the Method of Moments. Handbook of Mathematical Statistics*, edited by H. L. Rietz. 1924. (176)
- James, W. *Psychology, Briefer Course*. 1892. (26, 30, 31)
- Jennings, H. S. "Heredity and Environment." *The Scientific Monthly*. Sept., 1924. (39)
- Kelley, T. L. *Educational Guidance*. 1914. (60)
- Statistical Method*. 1923. (57, 97)
- McCall, W. A. *How to Measure in Education*. 1922. (97, 99)
- "A Proposed Uniform Method of Scale Construction." *Teachers College Record*, January, 1921. (140)
- Madsen, I. N. "Some Results with the Stanford Revision of the Binet-Simon Tests." *School and Society*, May, 1924. (196)
- Manny, F. A. and "The New Weight-Height Tables and Malnutrition." *Archives of Pediatrics*, October, 1924. (177, 187)
- Emerson, W. R. P.
- Miller, W. S. "The Variation of Significance of Intelligence Quotients Obtained from Group Tests." *Journal of Educational Psychology*, September, 1924. (196)
- Monroe, P. *Cyclopedia of Education*. 1911. (22)
- Monroe, W. S. *Introduction to the Theory of Educational Measurements*. (140)
- Measuring the Results of Teaching*. 1918. (60)
- Moore, E. C. *What Is Education?* 1915. (82)
- Murdoch, K. "The Measurement of Certain Elements in Hand Sewing." *Teachers College Contributions to Education*, 1919, No. 103. (104)
- Norsworthy, N. "The Psychology of Mentally Deficient Children." *Archives of Psychology*, No. 1. 1906. (24)
- O'Shea, M. V. and *The Child and His Spelling*. 1914. (143)
- Cook, W. A.
- Otis, A. *Statistical Method in Educational Measurement*. 1924. (88, 99)
- "The Reliability of Spelling Scales Involving a Derivation Formula for Correlation." *School and Society*, November 11, 1910. (143)
- "Some Logical Aspects of the Binet Scale." *The Psychological Review* Vol. XXIII, No. 2, Mar., No. 3, May, 1916. (148)
- Pritchett, H. S. *Seventeenth Annual Report of Carnegie Foundation for the Advancement of Teaching*. 1922. (12)
- Rankin, P. T. and "The Measurement of Composition Ability," an unpublished study. (136)
- Courtis, S. A.
- Rasey, M. I. "The Measurement of Spelling Ability," an unpublished study. (140)
- Rice, J. M. *Scientific Management in Education*. 1913. (12)
- Richards, A. M. and "Correlations of Single Measures of Some Representative Reading Tests." *School and Society*, 4:375. September, 1916. (88)
- Davidson, P. E.
- Rietz, R. S. *Handbook of Mathematical Statistics*. 1924. (176)
- Riley, J. L. *The Springfield Tests*. 1908. (12)
- Ruch, G. M. *The Improvement of the Written Examination*. 1924. (97)
- Rugg, H. O. *Statistical Method Applied to Education*. 1917. (57)

- Smith, E. R. Education Moves Ahead. 1924. (12)
- Spearman, C. The Nature of Intelligence and the Principles of Cognition. 1923. (26, 32, 83)
- Starch, D. Educational Measurements. 1916. (60)
- Stecher, L. I. and Baldwin, B. T. "Mental Growth Curves of Normal and Superior Children." University of Iowa Studies in Child Welfare, 1922. (189, 190, 195)
- Stern, W. Die Intelligenz der Kinder und Jungenlichen. 1920. (25)
- Stone, C. W. Arithmetical Abilities. 1908. (13)
- Terman, L. M. The Measurement of Intelligence. 1916. (25, 147)
- The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence. 1917. (148)
- Thiele, C. L. "The Measurement of Ability in Arithmetic," an unpublished study. (148)
- Thomson, G. H. and Brown, W. The Essentials of Mental Measurements. 1921. (83, 93, 94)
- Thorndike, E. L. Educational Psychology. 1913. (32, 35, 43, 45, 83, 84, 96, 113)
- "Handwriting." Teachers College Record, March, 1910. (104)
- "An Improved Scale for Measuring Ability in Reading." Teachers College Record, January, 1916. (88, 143)
- "Intelligence and Its Uses." Harper's Monthly Magazine, January, 1920. (46)
- "The Measurement of Achievement in Drawing." Teachers College Record, November, 1913. (104)
- "The Measurement of Intelligence." Psychological Review, 1924, 31:227-8, (84, 105, 124)
- , and
Curtis, S. A. "Correction Formulæ for Addition Tests." Teachers College Record January, 1923. (88, 90, 108, 124, 128)
- Tirrell, H. A. "The Norwich Tests." School Review, 1910, pages 326-337. (12)
- Trabue, M. R. Measuring Results in Education. 1924. (110)
- Ward, H. T. The New Social Order. 1919. (12)
- Whipple, G. M. "Endowment, Maturity and Training as Factors in Intelligence Scores." Scientific Monthly, 1924, 18:499, May. (45)
- Manual of Mental and Physical Tests. 1910. (23, 115)
- Williams, H. S. History of Science. 1904. (16, 22)
- Witmer, L. "Clinical Psychology." The Psychological Clinic, Vol. I, No. 1, 1907. (24)
- Woodrow, H. Brightness and Dullness in Children. 1919. (84)
- Wundt, W. Outlines of Psychology. 1897. (22)
- Psychologische, Psychologie. 1911. Vol. III. (32)
- Year Books National Society for the Study of Education
- "English Composition," Twenty-second, Part I (1923). (77-99)
- "Adapting the Schools to Individual Differences," Twenty-fourth, Part II (1925)
- Yerkes, R. M. and Yoakum, C. S. Army Mental Tests. 1920. (25)

